

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus_p2n model

Run on: January 31, 2003, 02:25:34 ; Search time 3121.29 Seconds
(without alignments)
4531.450 Million cell updates/sec

Title: US-09-825-414-7
Perfect score: 2412
Sequence: 1 MHINRRVQQPFPVTATDSFRT.....IEEGGTASPSEIFPRMRS 486

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL-frame+_p2n.model -DEV=xlp
-Q/cgn2.1/USPTO_spool/US09825414/runat_28012003_161741_26085/app_query.fasta_1.1294
-DB=GenEmbl -QFMT=fastap -SUFFIX=p2n.rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdl -LIST=45
-DOCCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09825414@cgn_1.1.4707_@runat_28012003_161741_26085 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_ln:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_ln:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*

29: em_vl:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	2412	100.0	52498	1	AF232004	AF232004 Pseudomon
2	1653	68.5	1834	1	AF458051	AF458051 Pseudomon
3	508	21.1	208050	1	AL646083	AL646083 Ralstonia
C 4	168	7.0	15311	1	AE007164	AE007164 Mycobacte
C 5	164	6.8	47852	1	MTV023	AL022022 Mycobacte
C 6	162.5	6.7	14131	1	AE005780	AE005780 Caulobact
7	160.5	6.7	42526	1	SC4A7	AL133423 Streptomy
8	156	6.5	10275	1	AE012164	AE012164 Xanthomon
9	155.5	6.4	13431	1	AE000234	AE000234 Escherich
10	155.5	6.4	15424	1	AE007019	AE007019 Mycobacte
11	155.5	6.4	16902	1	D90775	D90775 E.coli geno
12	155.5	6.4	18700	1	D90774	D90774 E.coli geno
13	151.5	6.3	5329	1	AF317649	AF317649 Rhodobact
14	151.5	6.3	203050	1	AL646078	AL646078 Ralstonia
15	151	6.3	40790	1	MTCY493	Z95844 Mycobacteri
16	150.5	6.2	6413	1	AF067776	AF067776 Abiotroph
17	149.5	6.2	348077	1	AP003000	AP003000 Mesorhizo
C 18	148.5	6.2	20701	1	AE006958	AE006958 Mycobacte
19	148.5	6.2	29550	1	MTV039	AL021942 Mycobacte
20	148	6.1	12412	3	AC006669	AC006669 Caenorhab
21	148	6.1	346897	1	AP002995	AP002995 Mesorhizo
22	147.5	6.1	6744	1	SSEPFAA	X71880 S.suis (189
23	147.5	6.1	6744	6	A24024	A24024 S.suis EF*
24	147.5	6.1	6744	6	I57310	I57310 Sequence 2
25	147.5	6.1	10052	1	AE004597	AE004597 Pseudomon
26	147	6.1	38404	1	SC2G5	AL035478 Streptomy
27	147	6.1	53662	1	MTV016	AL021841 Mycobacte
28	147	6.1	78210	1	AB070949	AB070949 Streptomy
29	147	6.1	188050	1	AL646072	AL646072 Ralstonia
30	146	6.1	15806	1	AE007151	AE007151 Mycobacte
31	145.5	6.0	10432	1	AE005333	AE005333 Escherich
32	145.5	6.0	18660	1	SCE2	AL049645 Streptomy
33	145.5	6.0	146500	2	AC119500	AC119500 leishmani
34	145.5	6.0	222605	1	AP002555	AP002555 Escherich
C 35	145	6.0	3701	1	AF172724	AF172724 Caulobact
C 36	145	6.0	11283	1	AE005710	AE005710 Caulobact
C 37	145	6.0	12070	1	AB011413	AB011413 Streptomy
38	145	6.0	110000	2	LMFLCHR18_07	Continuation (8 of
39	144	6.0	8126	1	AFANCC	L31363 Alcaligenes
40	144	6.0	189050	1	AL646077	AL646077 Ralstonia
41	143.5	5.9	6761	8	AB071164	AB071164 Saccharom
42	143.5	5.9	11830	1	BPEFHAB1	M60351 B.pertussis
43	143.5	5.9	11831	1	BPFHABFH	X52156 Bordetella
44	143.5	5.9	11883	6	I06606	I06606 Sequence 1
45	143.5	5.9	46819	6	AR204176	AR204176 Sequence

RESULT 1

ALIGNMENTS

AF232004/c
LOCUS 52498 bp DNA linear BCT 05-MAR-2001
DEFINITION Pseudomonas syringae pv. tomato strain DC3000 Hrp pathogenicity
island, complete sequence.
ACCESSION AF232004 AF061028 AF061029 AF232006 L41861
VERSION AF232004.3 GI:13325077
KEYWORDS
SOURCE
ORGANISM
Pseudomonas syringae pv. tomato.
Pseudomonas syringae pv. tomato
Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
Pseudomonas.
REFERENCE
AUTHORS 1 (bases 25494 to 29778)
TITLE Preston,G., Huang,H.C., He,S.Y. and Collmer,A.
The HrpZ proteins of Pseudomonas syringae pvs. syringae, glycinea,
and tomato are encoded by an operon containing Yersinia ysc
homologs and elicit the hypersensitive response in tomato but not
soybean
Mol. Plant Microbe Interact. 8 (5), 717-732 (1995)
JOURNAL
MEDLINE 96025089
PUBMED 7579616
REFERENCE
AUTHORS 2 (bases 22134 to 25847; 29687 to 32670)
TITLE Deng,W.L., Preston,G., Collmer,A., Chang,C.J. and Huang,H.C.
Characterization of the hrpC and hrpS operons of Pseudomonas
syringae pathovars syringae, tomato, and glycinea and analysis of
the ability of hrpF, hrpG, hrpC, hrpT, and hrpV mutants to elicit
the hypersensitive response and disease in plants
J. Bacteriol. 180 (17), 4523-4531 (1998)
JOURNAL
MEDLINE 98389667
PUBMED 9721291
REFERENCE
AUTHORS 3 (bases 31672 to 51723)
TITLE Charkowski,A.O., Alfano,J.R., Preston,G., Yuan,J., He,S.Y. and
Collmer,A.
The Pseudomonas syringae pv. tomato HrpW protein has domains
similar to harpins and pectate lyases and can elicit the plant
hypersensitive response and bind to pectate
J. Bacteriol. 180 (19), 5211-5217 (1998)
JOURNAL
MEDLINE 98422476
PUBMED 9748456
REFERENCE
AUTHORS 4 (bases 901 to 22404; 31672 to 51723)
TITLE Alfano,J.R., Charkowski,A.O., Deng,W.L., Badel,J.L.,
Petrnicki-Ocwieja,T., van Dijk,K. and Collmer,A.
The Pseudomonas syringae Hrp pathogenicity island has a tripartite
mosaic structure composed of a cluster of type III secretion genes
bounded by exchangeable effector and conserved effector loci that
contribute to parasitic fitness and pathogenicity in plants
Proc. Natl. Acad. Sci. U.S.A. 97 (9), 4856-4861 (2000)
JOURNAL
MEDLINE 20243785
PUBMED 10781092
REFERENCE
AUTHORS 5 (bases 1 to 52498)
TITLE Ramos,A.R., Rehm,A.H. and Collmer,A.R.
Pseudomonas syringae pv. tomato DC3000 hrpL through hrpC
unpublished
6 (bases 1 to 52498)
TITLE Alfano,J.R. and Collmer,A.
Direct Submission
Submitted (07-FEB-2000) Dept. Biol. Sci., UNLV, 1854 Maryland
Parkway, Las Vegas, NV 89154, USA
7 (bases 1 to 52498)
TITLE Ramos,A.R., Rehm,A.H. and Collmer,A.R.
Direct Submission
Submitted (22-NOV-2000) Plant Pathology, Cornell University, 334
Plant Sciences Bldg., Ithaca, NY 14850, USA
8 (bases 1 to 52498)
TITLE Sequence update by submitter
Ramos,A.R., Rehm,A.H. and Collmer,A.R.
Direct Submission
Submitted (05-MAR-2001) Plant Pathology, Cornell University, 334
Plant Sciences Bldg., Ithaca, NY 14850, USA
REMARK
AUTHORS
JOURNAL
TITLE
REMARK
COMMENT
FEATURES
Source

4/25/00 I have
too new.

/organism="Pseudomonas syringae pv. tomato"
/strain="DC3000"
/db_xref="taxon:323"
/note="exchangeable effector locus flanking the hrp/hrc
cluster
pathovar: tomato"
complement(908..1972)
/gene="queA"
complement(908..1972)
/gene="queA"
/codon_start=1
/transl_table=11
/product="QueA"
/protein_id="AAE71483.1"
/db_xref="GI:8037772"
/translation="MRVADFTFELPDSLIRAHPIAERRSSRLTLDPYTGALAHROFT
DLLEHLRSGDLMVFNNRVIIPARLFGQKASGKLEILVERVLDSHRVLAHVASKSPK
PGSSILIDGGGEAEVAVARDALFELRFAEVLPLDRVGHMPLPYIDRPDEGADRER
YQTVYAQRAGAVAAPTAGLHFDQPLMEAIAAKGVETAFVTLHVAGAGTFQPVRYQIED
HHMSEWLEVSQDVVDVAVACRARGVIAVGTTSVRSLESAARDGOLKPFSGTDIE
IYGRPEHVVDALVTNFIHLPSTLLMLVSAFAGYPTTAAAYAAIEHGYRFSYGDAM
FTRNPAPTAPQESAPEDHA"
2090..2177
/product="tRNA-Leu"
2343..2753
/note="ORF4"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAE71484.1"
/db_xref="GI:8037773"
/translation="MNKIYVKAYPEKPIGEVSVKYPTEIKKGFEGDKEIMKETQW
QOTGWSDCQIDGERLSKDVEDAVAQINADGYEIQVLPLSGAYDYALKRYELRHNK
TELSPGDQSYVFGYGYSTEGVTLVAKKFOSSAS"
complement(2908..3324)
/note="ORF3"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAE71485.1"
/db_xref="GI:8037774"
/translation="MRAVKNLJAKIGFLALTTIGSLPAFAVNDCLDNDNSTGAT
CGGNDKLDNDNVTDAFNGNDKMDNDHHTDAARGNDKDLONDHHTDAARGNDKD
LDNDNKTDAAFGGNDRLDNDNNTDNYNGTPSAAK"
complement(3939..4664)
/note="ORF2"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAE71486.1"
/db_xref="GI:8037775"
/translation="MOKTTLWALAFAMLAGCGVSGPAPGSDIOGAQAEKTPYKLNLD
AYTSKKLDVLEAFRTNKSXYMKGQLIDLVSGAFLGTPYRSNMLVGSANPEQLVIDER
GUDCFAYLDVFEAFRSTSQODEVRNLVQVRYKGDVDFLNKHFETDWAYGTAYPVA
DDITAQISPGAVSVKRLNERAKGVYLPGLPVVERSMYIIPRLVDSQVYSHLRTGD
YIGIYTPASRAGCDTRRFLYRDG"
5571..6410
/gene="tnpA"
5571..6410
/gene="tnpA"
/note="truncated at the amino terminus"
/codon_start=1
/transl_table=11
/product="tnpA"
/protein_id="AAE71487.1"
/db_xref="GI:8037776"
/translation="MGFLRPAYPLMAMLRHLMQNWEDYSDPAMEALYETTLRQFA
GLSLDRIADETTLNFRLEKHELAGLQVINGYLGRGLMLRQGMVVDATIIHAP
SSTKNKDKRDPKEMHQTKKNGYFFGMAHIGVDAESGLVHSLVGTAAVADTVQVDQ
LHSEETVYVSGDAGYTGVDKRAEHQDROMIWSIARPSRYKKHGEKSLIARYRKIEF
TKAQLRAKVEHPLRVIKRQFGYTKVRFRLAKNTAQATLFLASNLMMVRKRLAMGE
VRL"
complement(6656..8056)


```
/note="ORF1"
/codon_start=1
/translation_table=11
/product="unknown"
/protein_id="AAF71488.1"
/db_xref="GI:8037777"
/translation="MRPVGPAPGYPPTYEAEERPTAQAAGNDRARSSQASSSPASV
APETPMLGDLKRFPAGRYPMKMEVENRIKEQEPGKDGVKHTRRRKPPDAAGSSHVH
GGQSVASTSASQSKALQDTNFENKASDLAEARWCESPHRYALAPSKAAGKSSQLSANV
VSIHQEGHAEQRLEAQGLKADVVYSEGRDLHINLNYLEMDSCLGTSKGLWAPD
SNDKLLAKAARYFDDENAQKLPELAPLTKMKSKDSLGMRELLRAPGLVIGEGHNS
TSSKRELINMKSLLKASGVTTLFMEHLCAESHDKALNNYLSAPKGSMPARLKNYIDL
QSQGHQAPRELHTKYNFTTLVEAAKHAGLRVSLDTSTYMAPEKAELKRAQAMNYA
AEKIRLSKEGKWVAFVGATHTATSCDVGPLAELHGVRSVIDDLGLKSRAIVDINVK
NYGKLNPDVRLSYKV"
complement(8131. .10473)
/gene="hrpK"
complement(8131. .10473)
/gene="hrpK"
/codon_start=1
/translation_table=11
/product="HrpK"
/protein_id="AAF71489.1"
/db_xref="GI:8037778"
/translation="MRISSSPFVIVNQTPGELALAVESPLAKALETPVGGGQAGVQ
FGQAPAGNTQGAPTGAEOQTASSILALLQSGGAANANVNNVOPOPCADPAEAAPT
EATAPAPVEQVAAPTNTDATDASNAPKADAFLDNSEYSSPEALKRMEPMVAHLPP
EEREQAAKELNRPIDAAWMAADDSPDAGKAMDEFINAPLAKTAVDVAOGGNAWKIT
NKDLTFAKNMEKAADSADKDLANTYKDNPDADPQSLMEVRSAAVNRANMPLAIAADP
HHAVIDAADKTKVDGNVDAEDLKGASNPGLSKALKOSCTWSQPGCLGVDEAGMSG
RKKAAHSPDKMEFADAKNLSEWIKKSAPTINGQFASMLSDSATLNAVAGIDISLKDVE
DKPKSYSGAQKAAMVWKLOQTQOSVIAGRSLRNTDKTEQGLNDRIISQADAPDVQAYL
NKSIPEDQRNLVRSASLQKAVEQTKNVNSGQALQTDMDKADKAVKHNPNADYSGA
ISGLSAQLQLOKDLFPDSKVPPTNQVLENKRPDLQDKIATSYVTNFESEGALKQLLGQK
KSDASQSLQTDADNQKAIYDSVLPDFTENQHENYMNSTLGTLDQSKGRKLLGKSD
GGPSMLTQLAEAGVGKALNSIVGFASVSLIAHGDKLGATQAIYDSTRMGAELIKGG
IDAGAKMGREASAGLGRIGGOMIGRAVMVAGEATGLAAGALGAIPVVGWAIDGA
LAVGFGISMTIDAVKKHKAQKAFDHNVDVLDQFGIAKAH"
complement(10560. .11114)
/gene="hrpL"
complement(10560. .11114)
/gene="hrpL"
/note="alternative sigma factor"
/codon_start=1
/translation_table=11
/product="HrpL"
/protein_id="AAG33875.2"
/db_xref="GI:13325078"
/translation="MFQKIVILDSQPRPPSSAGIRQMTADQIQMRAFIQKRVMP
DDVDILQCYFLALRNEHKFQASHKPTWLCGIALNLIRNHFRRKMRQPYQESWED
VHSELEGHGDVSHQVEGHRQLARVIAIDCLPSNMOKVLEVSLEMGNVQETANSLGV
PIGTVRSRLSGARVOLKQIDPFA"
11358. .12464
/gene="hrpJ"
11358. .12464
```

Alignment Scores:

```
Pred. No.: 1.97e-122 Length: 52498
Score: 2412.00 Matches: 486
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 1 Gaps: 0
```

US-09-825-414-7 (1-486) x AEF232004 (1-52498)

```
QY 1 MetHsIlEaSnArGArVaIgInGInPrOpRoVaIthRaIaThraspSerPheaRgThr 20
Db 46117 ATGCACATCAACCGACGCGTCCAAACACCGCCTGTGACTGCGACGAGATAGCTTTCGAGCA 46058
QY 21 AlAsErAspAlAsErLeuAlAsErSerSerValArGSeRValSerSerAspGInGInArg 40
Db 46057 GCGTCGACGCGCTCTCTTGCCTCCAGCTCTGTGCGATCTGTCAAGCTCCGATCAGCAACGC 45998
```

```
QY 41 GluIIEaSnAlaIIEaIAspTYrLeuThraspHisValPheAlaAlaHisLysLeuPro 60
Db 45997 GAGATAAATGCGATTGCCGATTACCTGACAGATCATGTGTTCGCTGCCATAACTGCCG 45938
QY 61 ProAlAspSerAlAspGInGInAlaAlaValAspValHisAsnAGInIIEthraAla 80
Db 45937 CCGGCCGATTGCGCTGATGGCCAAAGCTGCAGTTGACGTACACAATGCGCAGATCACTGCC 45878
QY 81 LeuIIEgluThrArgAlAsErArgLeuHisPheGluGluGluThrProAlaThrIIEaIa 100
Db 45877 CTGATCGAGACGCGCGCCAGCCCTGCACCTTGAAGGGGAAACCCGGCAACCATCGCC 45818
QY 101 AspThrPheAlaLysAlaGluLysLeuAspArgLeuAlaThrThrThrSerGlyAlaLeu 120
Db 45817 GACACCTTCGCCAAGCGGAAAGCTCGACCGATTGGCGACGACTACATCAGCGCGCTTG 45758
QY 121 ArgAlaThrProPheAlaMetAlaSerLeuLeuGlnTYrMetGlnProAlaIIEasnLys 140
Db 45757 CCGGCCGACGCCCTTGGCCATGGCCCTCGTTGCTTCAGTACATGCAGCCTGCCATCAACAAG 45698
QY 141 GlyAspTrpLeuProAlaProLeuLysProLeuThrProLeuIIEserGlyAlaLeuSer 160
Db 45697 GCGGATTGGCTGCCGCTCCGCTCAAAACCGCTGACCCCGCTCATTTCCGAGCGCTGCG 45638
QY 161 GlyAlaMetAspGInValGlyThrLysMetMetAspArgAlaThrGlyAspLeuHisTYr 180
Db 45637 GCGGCCATGACCAAGGTGGGCAACAAGATGATGACCGCGCGAGCGGTGATCTGCATAC 45578
QY 181 LeuSerAlaSerProAspArgLeuHisAspAlaMetAlaAlaSerValLysArgHisSer 200
Db 45577 CTGAGCGCCTCGCCGACAGGCTCCACGATGCGATGGCCGCTTCGGTGAAGCGCCACTCG 45518
QY 201 ProSerLeuAlaArgGInValLeuAspThrGlyValAlaValGlnThrTYrSerAlaArg 220
Db 45517 CCAAGCCTTGTCTCGACAGGTTCTGCAACAGGGGGTTCGGGTTCAAGACGTACTCGCGCGCC 45458
QY 221 AsnAlaValArgThrValLeuAlaProAlaLeuAlaSerArgProAlaValGInGlyAla 240
Db 45457 AACGCCGTACGTACCGTATTGGCTCCGGCACTGGCGCTGCAGACCCGCGCTGCAGGGTGCT 45398
QY 241 ValAspLeuGlyValSerMetAlaGlyGlyLeuAlaAlaAsnAlaGlyPheGlyAsnArg 260
Db 45397 GTGACCTTGGTGATCGATGGCGGGGTGCTGCGTGCACACGACGCGCTTGCCAACCGCC 45338
QY 261 LeuLeuSerValGInSerArgAspHisGInArgGlyGlyAlaLeuValLeuGlyLeuLys 280
Db 45337 CTGCTCAGTGTGCAGTCGCGGTGATCACCAAGCGTGGCGGTGCATTAGTCTCGGTTGAAG 45278
QY 281 AspLysGluProLysAlaGlnLeuSerGluGluAsnAspTrpLeuGluAlaTYrLysAla 300
Db 45277 GATTAAGAAGCCCAAGGCTCACTGAGCGAAGAAACGACTGGCTCGAGGCTTATTAAGCA 45218
QY 301 IleLysSerAlaSerTYrSerGlyAlaAlaLeuAsnAlaGlyLysArgMetAlaGlyLeu 320
Db 45217 ATCAAAATCGGCCAGCTACTCGGGTCCGGCGCTCAACCGCTGCGCAAGCGGATGGCCGGTCTG 45158
QY 321 ProLeuAspMetAlaThraspAlaMetGlyAlaValArgSerLeuValSerAlaSerSer 340
Db 45157 CCACTGATATATGCGACCGACCGCAATGGGTGCGGTAAAGACCTGTGTCAAGCGTCCAGC 45098
QY 341 LeuThrGlnAsnGlyLeuAlaLeuAlaGlyGlyPheAlaGlyValGlyLysLeuGlnGlu 360
Db 45097 CTGACCCCAAAACGCTGCGCCTGGCGGGTGGCTTTCGAGGGGTAGGCAAGTTCAGAGAG 45038
QY 361 MetAlaThrLysAsnIIEthraspProAlaThrLysAlaAlaValSerGlnLeuThrasn 380
Db 45037 ATGGCGACGAAAAATATACACGACCGCGCGACCAAGCGCGGTCACTCAGTTGACCAAC 44978
QY 381 LeuAlaGlySerAlaAlaValPheAlaGlyTYrTrpThrThraAlaLeuThrThraspPro 400
Db 44977 CTGGCAGGTTCGGCAGCGGTTTTCGACAGGCTGCAGCACCGCGCGGTGACACACGATCCC 44918
```

Qy	401	AlavallyslsalagIuserPheIIegInasprThrVallysserThralaserSerThr	420
Db	44917	GCGGTGAAAAAGCCGAGTCTTCATACAGACACGGTGAAATCGACTGCATCCAGTACC	44858
Qy	421	ThrGlyTyrrValAlaaSpGlnThrVallyslLeuAlalystrVallylAspmetGlygly	440
Db	44857	ACAGGCTACGTAGCCGACGACCGCTCAAACTGGCGAAGACCGTCAAAGACATGGCGGG	44798
Qy	441	GluaIalleThrHisrHrclyAlaSerLeuArgasnThrValasnAnleuArgGlnArg	460
Db	44797	GAGGCGATCACCATACCGGCCACGCTTGCCGAATACGGTCAATAACTGCGTCAACGC	44738
Qy	461	ProAlaArgGlualaaSPIlegluGlyglyThrralaAlaSerProsergluilePro	480
Db	44737	CCGGCTCGTGAAGCTGATATAGAAGAGGGGGCAGCGGCTTCTCCAAGTGAATACCG	44678
Qy	481	pheaRgPrometarGser	486
Db	44677	TTTGCGCCTATGCGGTCG	44660
RESULT 2			
LOCUS	AF458051	1834 bp	DNA linear BCT 02-MAR-2002
DEFINITION	Pseudomonas syringae pv. maculicola type III effector HopToAIpma gene, complete cds.		
ACCESSION	AF458051		
VERSION	AF458051.1 GI:19071505		
KEYWORDS			
SOURCE	Pseudomonas syringae pv. maculicola.		
ORGANISM	Pseudomonas syringae pv. maculicola Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae; Pseudomonas.		
REFERENCE	1 (bases 1 to 1834) Guttman,D.S., Vinatzer,B.A., Sarkar,S.F., Ranall,M.V., Kettler,G. and Greenberg,J.T. A functional screen for the type III (Hrp) secretome of the plant pathogen Pseudomonas syringae Science 295 (5560), 1722-1726 (2002)		
JOURNAL MEDLINE	21862332		
PUBMED	11872842		
REFERENCE	2 (bases 1 to 1834) Vinatzer,B.A., Guttman,D.S., Sarkar,S.F., Ranall,M.V. and Greenberg,J.T. Direct Submission Submitted (13-DEC-2001) MCB, The University of Chicago, 1103 E 57th Street, Chicago, IL 60637, USA		
AUTHORS	location/Qualifiers		
TITLE	1..1834		
JOURNAL	/organism="Pseudomonas syringae pv. maculicola"		
FEATURES	/strain="ES4326"		
source	/db_xref="taxon:59511"		
	/note="pathovar: maculicola"		
	91..116		
	/note="hrp box"		
CDS	280..1734		
	misc_feature		
	/note="similar to Pseudomonas syringae pv. tomato HopToAI encoded by GenBank Accession Number AF232004"		
	/codon_start=1		
	/transl_table=11		
	/product="type III effector HopToAIpma"		
	/protein_id="AAL84253.1"		
	/db_xref="GI:19071506"		
	/translation="MYIKRSISSOSIGTESFHSAQSVASSVRVLESEAOTKLNDIT HYLDHYFAHRLPVNDINDODVVHAHNVEIGKIISARALRLDEGETALSIGETFA KAKEFDPMASASGALRATPEAASYOYMPAINKGDMLPTPLKPLAPLVSGALSGV MDQVGTVGNRRATGDLHYLSTAPEKLHDMAASVKRHKGVMRQAADLGIAVQTYTAR NALRTVLAPALASRPAYOGAVDISAAGSLVANAGFSERMHVTVOARDHLRGAFVLG IKDKQPKRADLSEETDWDLAYRAIKSASYSAGALNAGKRVAGLPLDAVTDGLKAVRSIV SATSLMONGILVAGGFAGVGKIQEMATKNITHPATKAAYSQTLNLASAPVFSAWTTA AVATDPVKKAESEFIQDTVKTTSNGTSALADKTVKLAKAGIDASAERIATTGASLRD TLRRITYREPDIIEGGVAAGTSPVAVPFEEAMRP"		
BASE COUNT	396 a	536 c	540 g 362 t
ORIGIN			

Alignment Scores:		1.52e-82	Length: 1834
Pred. No.:	Score:	1653.00	Matches: 337
Percent Similarity:	Best Local Similarity:	79.30%	Conservative: 50
Query Match:		69.06%	Mismatches: 93
DB:		68.53%	Indels: 8
		1	Gaps: 3
US-09-825-414-7 (1-486) x AF458051 (1-1834)			
QY	1 MethisIleAsnArgArgValGInGlnProProValThraIeThrasSerPheArgThr	20	
Db	280 ATGTACATCAACCGCTCCATCAGTTCTCAGTCCAGTATTGGCAGAGAGCTTCATTCT	339	
QY	21 AlaSerAspAlaSerLeuAlaSerSerSerValArgSerValSerSeraSpGInGlnArg	40	
Db	340 GCGCAG-----TCTGTGCTCCCTCAGTTCCGTCAGGGTGCTCAGTGAGCCGAGCAGACC	393	
QY	41 GluIleAsnAlaIleAlaAspTyrLeuThrasPHisValPheAlaAlaHisLysLeuPro	60	
Db	394 AACTGAATGACATCACCCCATTTACCTGACAGATCATGTATTCGCCGACACAGGTTACCG	453	
QY	61 ProAlaAspSerAlaAspGlyGlnAlaAlaValAlaSpValHisAsnAlaGlnIleThraAla	80	
Db	454 GTAACGATATATCAACGTTGATCATGAGATGTCGTTTCATGCCCCACACAGCTCGAGATCGCAAA	513	
QY	81 LeuIleGluThrArgAlaSerArgLeuHisPheGluGlyGluThrProAlaThrIleAla	100	
Db	514 ATATATCAGCGCCCGTGCCCTACGCTTGCTTGATGAGGGCGCAACCCGCTCTCAGCATTTGGC	573	
QY	101 AspThrPheAlaIalysAlaGluLysLeuAspArgLeuAlaThrThrSerGlyAlaLeu	120	
Db	574 GAAACCTTTGCCAAGGAGAGAGATTTCAGCCGATGGCAAGTGCGCCCTCGGGCGCTTAA	633	
QY	121 ArgAlaThrProPheAlaMetAlaSerLeuLeuGlnTyrMetGlnProAlaIleAsnLys	140	
Db	634 AGGGCAACACCGTTGCGCGGGCGTCCGTGCTGCAATACATGACGCTGCCATCAACAAG	693	
QY	141 GlyAspTrpLeuProAlaProLeuLysProLeuThrProLeuIleSerGlyAlaLeuSer	160	
Db	694 GCGCAGCTGTTACCGCAGCGCCGCTCAAGCCCTTGCCCCGCTCGTGTCCGGCGCCCTATCG	753	
QY	161 GlyAlaMetaspGlnValGlyThrLysMetMetaspArgAlaThrGlyAspLeuHisTyr	180	
Db	754 GCGCTCATGGACACAGGTGCGGCGCGCTCATGAAACCGCGGCGGACCTGCATTAC	813	
QY	181 LeuSerAlaSerProAspArgLeuHisAspAlaMetAlaAlaSerValLysArgHisSer	200	
Db	814 CTGAGCACTGCGCGCGAATAACTGCACGATGCGATGGCTGCTCGTAAACGCCCATAAAG	873	
QY	201 ProSerLeuAlaArgGlnValLeuAspThrGlyValAlaValGlnThrTyrSerAlaArg	220	
Db	874 CCGGCTGTCATGCGTACGCGCTGCCGATCTGGGGATCGCCGTTCAAGCTACACGGCGGA	933	
QY	221 AsnAlaValArgThrValLeuAlaProAlaLeuAlaSerArgProAlaValGlnGlyAla	240	
Db	934 AATGCGCTGCGTACCGTGTGCTCTGCTGCACTGGCCTCCAGACTGCAGTGACAGGGTGGC	993	
QY	241 ValaspLeuGlyValSerMetAlaGlyGlyLeuAlaAlaAsnAlaGlyPheGlyAsnArg	260	
Db	994 GTAGACATCAGCGTGTGCGCGGGGGTAGCCTGTTGCCAACGAGGTTTAAAGCAGCGC	1053	
QY	261 LeuLeuSerValGlnSerArgAspHisGlnArgGlyGlyAlaLeuValLeuGlyLeuLys	280	
Db	1054 ATGCATACGGGTGACAGGCGCGTGTACCTGCGTGGCGGTGCTTTGTACTCGGCATCAAG	1113	
QY	281 AspLysGluProLysAlaGlnLeuSerGluGluAsnAspTrpLeuGluAlaTyrLysAla	300	
Db	1114 GACAAACAACCCCAAGGACAGACTTGAGTGAAGAAACCGACTGCTCGATGCTTACCGGGCG	1173	
QY	301 IleLysSerAlaSerTyrSerGlyAlaAlaLeuAsnAlaGlyLysArgMetAlaGlyLeu	320	

Db	1174	ATCAAAATCCGCCACGACTACTCCGGGTGCAGCCGCTCAACGCTGGCAAGCGTGTCCCGGGTCTG	1233
QY	321	PROLEuAspMeTaLaThraspAlaMetGlyAlaValArgSerLeuValSerAlaSerSer	340
Db	1234	CCGCTGGACGCTGCCACCGATGGTTGAAAGCGGTGAGGAGTCTGGTGTCCGCCACGACC	1293
QY	341	LeuThrGlnAsnGlyLeuAlaLeuAlaGlyGlyPheAlaGlyValGlyLysLeuGlnGlu	360
Db	1294	CTGATGCAAAACGGTCTGTGTAATGGCGGTGGTTTGCCTGTGGTAAAGTTGCAGGAA	1353
QY	361	MeTaLaThrLysAsnIleThraspProAlaThrLysAlaAlaValSerGlnLeuThrasn	380
Db	1354	ATGGCGACGAAATAATATCACCCACCGGCAACCAAGCAGCCGTACAGCAACTGACCAAT	1413
QY	381	LeuAlaGlySerAlaAlaValPheAlaGlyTyrThrThrAlaAlaLeuThrThraspPro	400
Db	1414	CTTGACAGTTCAGGCCCGGTTTTCAGCCTGGACACGCGGAGCGGAGGACCA	1473
QY	401	AlaValLysLysAlaGluSerPheIleGlnAspThrValLysSerThrAlaSerSerThr	420
Db	1474	GCGGTGAAAAAGCGTGTGCTTCTTGACAGACACGCTGAACCAACAGGTATCCAACGGG	1533
QY	421	ThrGlyTyrValAlaAspGlnThrValLysLeuAlaLysThrValLysAspMetGlyGly	440
Db	1534	ACCAGCGCGCTGGCCGACAAACCGTCAAGCTGGCCAGGCGGCAATTGACGCGAGCGCC	1593
QY	441	GluAlaIleThrHisThrGlyAlaSerLeuArgAsnThrValAsnAsnLeuArgGlnArg	460
Db	1594	GAAAGATGGCCACTACCGGGCGCTCTTGGCGTGACACC-----TTGCGCGCGTGC	1644
QY	461	ProAlaArgGluAlaAspIleGluGluGlyGlyThrAlaAla-----SerProSer	477
Db	1645	ACGGTGCCTGAGCCTGATATCGAGAGAGCGCGGTGTGCGCGGACAGGGTTCCCGGTT	1704
QY	478	GluIleProPheArgPrometArg	485
Db	1705	GCAGTGCCTTTCGAGCGCAATGCGG	1728
RESULT 3			
AL646083			
LOCUS	AL646083	208050 bp	DNA linear BCT 07-DEC-2001
DEFINITION	Ralstonia solanacearum GM1000 megaplasmid, complete sequence;		
ACCESSION	AL646083		
VERSION	AL646083.1	GI:17431588	
KEYWORDS			
SOURCE	Ralstonia solanacearum.		
ORGANISM	Ralstonia solanacearum.		
REFERENCE			
AUTHORS	1 (bases 1 to 208050)		
TITLE	Salanoubat,M., Genin,S., Artiguenave,F., Gouzy,J., Mangenot,S.,		
REFERENCE	Arlat,M., Billault,A., Brottier,P., Camus,J.C., Cattolico,L.,		
AUTHORS	Chandler,M., Choisine,N., Claudel-Renard,C., Cunnac,S., Demange,N.,		
TITLE	Gaspin,C., Lavie,M., Moisan,A., Robert,C., Saurin,W., Schlex,T.,		
REFERENCE	Sigulier,P., Thebault,P., Whalen,M., Wincker,P., Levy,M.,		
AUTHORS	Weissenbach,J. and Boucher,C.A.		
JOURNAL	Genome sequence of the plant pathogen Ralstonia solanacearum		
REFERENCE	Unpublished		
AUTHORS	2 (bases 1 to 208050)		
TITLE	Direct Submission		
JOURNAL	Submitted (05-DEC-2001) Genoscope and CNRS UMR-8030, 2 rue Gaston		
REFERENCE	Cremieux, CP5706, 91057 Evry Cedex, France, Laboratoire de Biologie		
AUTHORS	Moléculaire des Interactions Plantes-Microorganismes INRA-CNRS,		
TITLE	BP27, 31326 Castanet-Tolosan Cedex, France, Fondation Jean		
JOURNAL	Dausset-CEPH, 27 rue Juliette Dodu, 75010 Paris, France, LMGM CNRS		
REFERENCE	118 Route de Narbonne, F 31062 Toulouse Cedex 4, Genoscope and INRA		
AUTHORS	URGV, 2 rue Gaston Cremieux, CP5706, 91057 Evry Cedex, France,		
TITLE	Laboratoire de Biométrie et Intelligence Artificielle INRA, BP27,		
REFERENCE	F31326 Castanet-Tolosan Cedex. Laboratoire de Génétique Cellulaire		
AUTHORS	INRA, BP27, F31326 Castanet-Tolosan Cedex		
JOURNAL	Christian.Boucher@toulouse.inra.fr		

```

http://sequence.toulouse.inra.fr/R.solanacearum.html.
FEATURES
source
1..208050
/organism="Ralstonia solanacearum"
/strain="GM11000"
/db_xref="taxon:305"
/plasmid="megaplasmid"
135..947
/gene="RSp1115"
/note="RS05505"
135..947
/gene="RSp1115"
/function="miscellaneous; not classified regulator"
/note="Product confidence : putative
Gene name confidence : hypothetical
predicted by Homology
predicted by Framed"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="PUTATIVE TRANSCRIPTION REGULATOR PROTEIN"
/protein_id="CAD18266.1"
/db_xref="GI:17431589"
/translacion="MADSVQYQVADVPCGLVLSAGRFQSFERRHFLDFHVGIVTAG
VQRQRFKGTIVLGPCCISILMPPEIHGVAEGDGYTLKTFRLSQLLANLAQDISG
TDHEPIAATLLEDPALAGHLIHLHDAMHSGASTLOYOSEWLLALFLKOSRALIP
EHTKGALSPAQWQRYRDYCFAHAGERITLDDLASLCALGRFQFLKQTVGMTPHAW
LVRLRLQACSLSSRSKAIAEVAQEVGFYDQSHFNRAFRQAFGVAPSHYQS"
1162..2259
/gene="RSp1116"
/note="RS05504"
1162..2259
/gene="RSp1116"
/EC_number="4.1.2.5"
/function="small molecule metabolism; degradation; amino
acids"
/note="Product confidence : putative
Gene name confidence : hypothetical
predicted by Codon_usage
predicted by Homology
predicted by Framed"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="PUTATIVE LOW SPECIFICITY L-THREONINE ALDOLASE
PROTEIN"
/protein_id="CAD18267.1"
/db_xref="GI:17431590"
/translacion="MNRDPMPTTRTPALGFTSDNIEGASPEVLDAILANAGQAPYG
ADDCTARVEQTLSELEFEREVKFLVPTGTANALCLGAMTPPWGNYCHPSSHINDE
CGAPEFYTGAKLVAVDGPSAKIDVEKLRATRVKVGVDVHSTQPCVSLQATEVGS
YLNEIEAIGQLCDARIKLQMDGSRFANALVSLGCSPAAMTWKAGDALSFGATKNG
VVAEAIVLFDLSLAEMGRKRRAGHLFSKMRFLSAQIEAYLANDMLRNAROANDM
ALRLARGLEGMHGASVLGATEANIVFCRLPSTLLENLLQAGFRFYHNRMEPNVRFVT
SFATTADVDVNLNRHMKAAAGPSADTHATA"
2257..5019
/note="ACUR or RS05503
predicted by Framed"
/evidence=not_experimental
complement(2462..3139)
/gene="RSp1117"
/note="RS05502"
complement(2462..3139)
/gene="RSp1117"
/function="miscellaneous; unknown"
/note="Product confidence : putative
Gene name confidence : hypothetical
predicted by Codon_usage
predicted by Framed"
/codon_start=1
/evidence=not_experimental
/transl_table=11

```

```

/ product="PUTATIVE TRANSMEMBRANE PROTEIN"
/ protein_id="CAD18268.1"
/db_xref="GI:17431591"
/ translation="MNDASLHEHTIISCFLEHORPPTVREIAERFNDEAQRALRALI
ADNHGVLPHSDEVWIAHPESAFTTCVKSNGNRKWWNCAMCALGLVHLAGSATI
ETRIGAIDDHVIIRIENGRLLDTDYVHFPIPMKHAMDNVITYCSVQLLFRDEAQVDE
WCASRGIPKGDVRLPIEQVWNVFAVEWYARHADADWTKWSVREALIELFARHHLSGPVWTI
AAEAARF"
complement(3188. .3595)
/ gene="RSp1118"
/ note="RS05501"
complement(3188. .3595)
/ gene="RSp1118"
/ function="miscellaneous; hypothetical/global homology"
/ note="Product confidence : hypothetical
Gene name confidence : hypothetical
predicted by Codon_usage
predicted by Homology
predicted by Framed"
/ codon_start=1
/ evidence=not_experimental
/ transl_table=11
/ product="CONSERVED HYPOTHETICAL PROTEIN"
/ protein_id="CAD18269.1"
/db_xref="GI:17431592"
/ translation="MFDTKIAIVVRDDLATWOKLNTAFILMSGITGASPEIMGEAYRD
AADNTYHSLCQPVIVLSCDQETLKAVHRALERKVPGLYIEDMFSIGHDANRAVF
AEHGPD SARVVGIALREKKIVDKITGARHP"
complement(3673. .4161)
/ gene="RSp1119"
/ note="RS05500"
complement(3673. .4161)
/ gene="RSp1119"
/ function="miscellaneous; hypothetical/global homology"
/ note="Product confidence : hypothetical
Gene name confidence : hypothetical
predicted by Codon_usage
predicted by Homology
predicted by Framed"
/ codon_start=1
/ evidence=not_experimental
/ transl_table=11
/ product="CONSERVED HYPOTHETICAL PROTEIN"
/ protein_id="CAD18270.1"
/db_xref="GI:17431593"
/ translation="MGWSEVGETVCPARTLAVVGDRTVLILREILGVKRFEEFOA
QTGMSSHLSTRKRLADGIVVRHLYSDRPRHEXYRLTSGLDYFLLLSKSGEKG
WGFKAKTAPALITTRQCGHTGLKVCAPACDEPFGRKDATVTILGASFKAERQARD
GA"
4228. .4725
/ gene="RSp1120"
/ note="RS05499"
4228. .4725
/ gene="RSp1120"
/ function="miscellaneous; hypothetical/global homology"
/ note="Product confidence : hypothetical
Gene name confidence : hypothetical
predicted by Codon_usage
predicted by Homology
predicted by Framed"
/ codon_start=1
/ evidence=not_experimental
/ transl_table=11
/ product="CONSERVED HYPOTHETICAL PROTEIN"
/ protein_id="CAD18271.1"
/db_xref="GI:17431594"
/ translation="MPAITRHPRCAPRYLLEDFKVGDTLETSGFAVTRDMILSFAEY
DQPMHLDETAARDTVFGLGSGWQLAVTMRLLDARLIGTPIYGAEIRDIREHA
PMRPGDVLKRAAEVIAIRPSKSRPDRGFM DARVTTTNSEGVTLVIGRWSLVVPRRPF
OKVAF"
complement(5017. .5610)
/ gene="RSp1121"
/ note="RS05498"

```

CDS	complement(5017..5610) /gene="RspJ1121" /function="miscellaneous; not classified regulator" /note="Product confidence : putative Gene name confidence : hypothetical predicted by Codon_usage predicted by Homology predicted by Framed" /codon_start=1 /evidence=not_experimental /transl_table=11 /product="PUTATIVE TRANSCRIPTION REGULATOR PROTEIN" /protein_id="CAD18272.1" /db_xref="GI:17431595" /translation="MAEKITRKPRADAERNRQLLDVAKTAFEKVSASLEDIARE AGVGIGTLVRHPTRDALIDEIYRDEGSRLAEAROLSEDPLEAVRQMLLFVGL ANKQIRADVLCNMTDCSERLCTLSGEVLTETLAQLIERAKQSGAIGLAVEPLDLLSAV AGVASFGALDMEAGARRLVEVMVAGLRVGAATGGRK" 5822..6910 /gene="RspJ1122" /note="RS05497"		
gene			
Alignment Scores:			
Pred. NO.:	6.4e-17	Length:	208050
Score:	508.00	Matches:	171
Percent Similarity:	46.73%	Conservative:	72
Best Local Similarity:	32.88%	Mismatches:	217
Query Match:	21.06%	Indels:	60
DB:	1	Gaps:	14
US-09-825-414-7 (1-486) x AL646083 (1-208050)			
QY	3	ILEASNAARGARGVALGINGLNPROPROVALTHRALATHRASPSERPHARGTHRALASER	22
Db	203441	ATTCCGCCGACGCTCTCCGCCATCCCGAGGGCGCTGCTGAGCGGCCAATCCGAGCCGACGA	203500
QY	23	ASPLASERLEUALASERSERVALARGSERVALSERASPSGLNARG-GLU	42
Db	203501	ACAGCGATGTGAGCCAGTGTCTCCGCGCGGCGCTCCATCCGCGCACAAATCGATCGGAT	203560
QY	42	EASNALATLEALASPTYLTLEUTHRASPHISVALPHEALALAHISLYSLEUPROVAL	62
Db	203561	CCATTCGAATC--GAATTCGCGCGCGGGGGGCTCACCCGCCGATACCGCCGACCCCGGGC	203617
QY	62	AASP-----SERALASPSGLNALALALVAL-ASPY	73
Db	203618	AGACCGGCGCGTGCGATCGACATCCCCACTGGATGCCGATAGCCGACACAGCTGCGCGATA	203677
QY	73	AL-----HISASNALAGNLLETHRALALEUILEGLUTHRA	85
Db	203678	TTTTTCAGGCGCGCACTTTCAGCGGTACCAACGATGCGCAATTCGAGATCCTGATCCAGCAGC	203737
QY	85	RGALASERARGLEUHHISPHGLUGLGLUTHRPROALATHRILEALASPTHRPHLEAL	105
Db	203738	GCGCAGAACGGCTCCCAAGCGATGGGGGAACGCGCGGCAACCGTCCGCGCTACTGGCCA	203797
QY	105	YSALAGLULYSLEUASPARGLEUALATHRTHRTHRSERGLYALALEUARGALATHRPROP	125
Db	203798	AGGCGCCCAACCGGAGCCGGCTGGGCCAGACACAGGTAGGCTTCTGCGCTGCGTGGCGGT	203857
QY	125	HEALAMETALASERLEUENLTYRMTGLNPROALILEASNLYSGLYASPTRYLEUP	145
Db	203858	TCGCGCATTTGCCCTCGCGCCTGTTGCAGCGTGAAGCAGGCACTGACGGCCTTC-----	203907
QY	145	ROALAPROLEULSPROLEUTHRPROLEUILESERGLYALALEUSERGLYALAMETASPC	165
Db	203908	--GCCAAGACGACGCGCCCGAGTCGCGCGGACGCGTCCGGTCCGGGTCGGGACGCGCGATG	203965
QY	165	INVALGLYTHRLYSMETMETSPARGALATHRGLYASPLEUHHISTYRLEUSERALASERP	185
Db	203966	CCTTCGGGGGGCACCCCTGCTGGGCAAGGCCACACGACACAAATGGCTGGCGGCGCTCGC	204025
QY	185	ROASPARGLEUHHISASPALAMETALALASERVALLYSARGHISSEPROSERLEUALAA	205

Db 204026 CCGATCACCTCGAACCCTGTCATGGCGGCAAGCCGCAATAGGCCGTGCAGCCCACTCTGGGCC 204085
 QY 205 rgGlnValLeuAspThrGlyValAlaValGlnThrTyrSerAlaArgAsnAlaValArgT 225
 Db 204086 GGCTGGCTGGCGGAGGTGAGCCTGGCATTCAGACTATTCCCTGGCGCAACGTGATCCGCA 204145
 QY 225 hrValLeuAlaPro-----AlaLeuAlaSerArgProAlaValGlnGlyAlaVal 241
 Db 204146 CCGGGCTGGCGCCGCTGGCAACCCATGCCCTGGCGCGGCAAGGGCAGCCAAC-----G 204199
 QY 241 aLAspLeuGlyValSerMetAlaGlyGlyLeuAlaAlaAsnAlaGlyPheGlyAsnArgL 261
 Db 204200 TGGATTGCTGATCGCGGCGGTGGCGGCCGCTGGCCGCGCTACATGGCCA 204259
 QY 261 euLeuSerValGlnSerArgAspHisGlnArgGlyGlyAlaLeuValLeuGlyLeuLysA 281
 Db 204260 TGCAGCACATGAATGAGACCCACCACCGCACCGGGGGGGAATACCTGCTGGGC----- 204312
 QY 281 sPlysGluProLysAlaGlnLeuSerGluGluAsnAspTrpLeuGluAlaTyrLysAlaI 301
 Db 204313 -----CGCACCGACTGGGAAGAGACCAGTTCAGCAGC 204343
 QY 301 lElysserAlaSerTyrSerGlyAlaAlaLeuAsnAlaGlyLysArgMetAlaGlyLeuP 321
 Db 204344 TCAAGCAATCGACGTGACTGACCCGCTGGTGGGAGCGGGCAAGCGCACGCAAGCTGC 204403
 QY 321 roLeu--AspMetAlaThrAspAlaMetGlyAlaValArgSerLeuValSerAlaSerS 340
 Db 204404 TCGTCGCCGATTGGCGACCGAGACGCTGGCGCCGACGCGCAGCCTGTTCACCGCCACCA 204463
 QY 340 eRLeuThrGlnAenGlyLeuAlaLeuAlaGlyGlyPheAlaGlyValGlyLysLeuGlnG 360
 Db 204464 ACCTCATCAAGAACATGGGGCGCACTGGCGGGGTGGCTTGGCCGGTGTGACCGCACAGA 204523
 QY 360 lMetAlaThrLysAsnIleThrAsp-----ProAlaThrLysAlaAlaValS 376
 Db 204524 CCGCTCGCGGCAAGCGCCGACGAGCGCGCTATACCGAGCGCGCGCTCGCGCCGCTCA 204583
 QY 376 eRglnLeuThrAsnLeuAlaGlySerAlaAlaValPheAlaGlyTrpThrThrAlaAlaL 396
 Db 204584 GCGCGCGCGTCAAGCAGGTACTGTCCGACCGCGTCTACGCTGCGTGACACCGCCGATG 204643
 QY 396 euThrThrAspProAlaValLysLysAlaGluSerPheIleGlnAsp-----ThrValL 414
 Db 204644 TCATGGCGCGGCCGCCCATCGACGCCGCGCGCGGCACATCCAGCAAGCTTCACACCGCC 204703
 QY 414 ySSerThrAlaSerSerThrThrGlyTyrValAlaAspGlnThrValLysLeuAlaLysT 434
 Db 204704 TGATGCCGATGCCGCGCGCCCGCGGCACAGCGCATGACATGCTTAACGCGCAAAAGC 204763
 QY 434 hrValLysAspMetGlyGlyGluAlaIleThrHisThrGlyAlaSerLeuArgAsn--T 453
 Db 204764 CGGCGAGCAGGCTACTCGGTCTCGACCTC---CATCCGCTGACCGTCTGGGAGCCGCC 204820
 QY 453 hrValAsnAsnLeuArgGlnArgProAlaArg-----GluAlaAspI 467
 Db 204821 TCGTGGCGCAGCTGGCGGAAGATCCACCCCGCGCGGTGCTGATGAGGCCCAAGACAGATG 204880
 QY 467 leGluGluGlyGlyThrAlaAlaSerProSerGluIleProPheArgProMetArg 485
 Db 204881 TA-----GGCCTTGTGAAAGCGCCCGCAGCGTATCGCGCGGCTGCCCGCCGCC 204927

RESULT 4
 AE007164/c AE007164 15311 bp DNA linear BCT 27-APR-2001
 LOCUS Mycobacterium tuberculosis CDC1551, section 250 of 280 of the
 DEFINITION complete genome.
 ACCESSION AE007164 AE000516
 VERSION AE007164.1 GI:13883468
 KEYWORDS
 SOURCE Mycobacterium tuberculosis CDC1551.
 ORGANISM Mycobacterium tuberculosis CDC1551

REFERENCE
 AUTHORS
 Fleischmann,R.D., Alland,D., Eisen,J.A., Carpenter,L., White,O.,
 Peterson,J., Deboy,R., Dodson,R., Gwin,M., Haft,D., Hickey,E.,
 Kolonay,J.F., Nelson,W.C., Umayam,L.A., Ermolaeva,M.,
 Salzberg,S.L., Delcher,A., Utterback,T., Weidman,J., Khourl,H.,
 Gill,J., Mikula,A. and Bishai,W.
 TITLE
 Whole genome comparison of Mycobacterium tuberculosis clinical and
 laboratory strains
 JOURNAL
 REFERENCE
 AUTHORS
 2 (bases 1 to 15311)
 Fleischmann,R.D., Alland,D., Eisen,J.A., Carpenter,L., White,O.,
 Peterson,J., Deboy,R., Dodson,R., Gwin,M., Haft,D., Hickey,E.,
 Kolonay,J.F., Nelson,W.C., Umayam,L.A., Ermolaeva,M.,
 Salzberg,S.L., Delcher,A., Utterback,T., Weidman,J., Khourl,H.,
 Gill,J., Mikula,A. and Bishai,W.
 TITLE
 Direct Submission
 JOURNAL
 Submitted (25-APR-2001) The Institute for Genomic Research, 9712
 Medical Center Dr, Rockville, MD 20850, USA
 FEATURES
 source
 1. 15311
 /organism="Mycobacterium tuberculosis CDC1551"
 /strain="CDC1551"
 /db_xref="taxon:83331"
 /note="clinical strain"
 complement(114. 1662)
 /gene="MT3613"
 /note="This region contains an authentic frame shift and
 is not the result of a sequencing artifact; similar to
 GI:U00020 PID:467119; identified by sequence similarity;
 putative; acetolactate synthase, large subunit, putative,
 authentic frameshift"
 complement(1659. 2495)
 /gene="MT3614"
 complement(1659. 2495)
 /gene="MT3614"
 /note="identified by Glimmer2; putative"
 /codon_start=1
 /product="conserved hypothetical protein"
 /protein_id="AAK4792.1"
 /db_xref="GI:13883469"
 /translation="MTIDVWMOHPYQRLHGDMPASLRWTGGSIPETDPIEATVSS
 MDAGEVTGLLSAWRGPNGODLISNDVAWEVRLYPNFRAGLAAVDLDRPMAVRELR
 RRVGEFVGLRVVPMWLGAPPTDRRYPLFAECVQSAVPCTQVGHGTPLRPSETGRP
 IPIDOVALDEPVLIVCGHVGYPWTEEMVAVARKHENYIDTSAYTIKRLPGKLVRF
 MPTDTGQRKVLFGTNPYMIATHALTGLDELGLSDEARDELHGNVAVREKLDPRKY
 QT"
 2855. 8554
 /gene="MT3615.1"
 /note="This region contains an authentic frame shift and
 is not the result of a sequencing artifact; similar to
 GI:AL123456; identified by sequence similarity; putative;
 PE_PGRS family protein"
 complement(8683. 9339)
 /gene="MT3615.2"
 complement(8683. 9339)
 /gene="MT3615.2"
 /note="similar to GP:4455741; identified by sequence
 similarity; putative"
 /codon_start=1
 /transl_table=11
 /product="fatty-acid-CoA ligase-related protein"
 /protein_id="AAK4793.1"
 /db_xref="GI:13883470"
 /translation="MAASLSENLSCSSNMCRLSGNAATNLERGPGEPPGDRCTRQGA
 VRPARTLAKGNIPVGYXDEKRTAETFTINGVRVATFGDYAOVEEDGTVTMLGGS
 VSI NSGGEKVYPPEVEALAKHPDVFALVGVPPRRGOQVAAYVQARPGCRPSLAE
 LDSFVRSEIAGYKVPRLWFVDEVKRSAPGKPDYRWAKQTEARPADDVHAGHYTSGS
 "
 9385. 13038

gene

CDS	/gene="MT3615.3"	9385. .13038	/gene="MT3615.3"	/note="similar to GB:AL123456; identified by sequence similarity; putative"	/codon_start=1	/transl_table=11	/product="PE_PGRS family protein"	/protein_id="AAK47974.1"	/db_xref="GI:13883471"				
gene	complement(13384. .13602)	/gene="MT3615.4"	complement(13384. .13602)	/note="identified by Glimmer2; putative"	/codon_start=1	/transl_table=11	/product="--substrate--CoA ligase, putative"	/protein_id="AAK47976.1"	/db_xref="GI:13883473"				
CDS	/transl_table=11	/product="hypothetical protein"	/protein_id="AAK47975.1"	/db_xref="GI:13883472"	/translation="MMGAATSGHPARKVDMGRSLSLSGAEPSTMMMEVHRHGCSRAS RNSCGFAEAGPHLIDGHRALSGVAIG"	complement(13599. .15245)	/gene="MT3616"	complement(13599. .15245)	/gene="MT3616"				
gene	/note="similar to GP:2645200; identified by sequence similarity; putative"	/codon_start=1	/transl_table=11	/product="--substrate--CoA ligase, putative"	/protein_id="AAK47976.1"	/db_xref="GI:13883473"	/translation="MAVALNIDIAEHAIDAVPDRVAVICGEOLTYAOLEDKANRLA HHLIDQGVQKDKVGLYCRNRNRIEYIAMLGIVKAGAILVNVNFRYEGELRYLFDNSD MVALVHERRYADRVANVLPDTPHYRTILVEDGSDODRYRGVGFERYSAIAAGSPERD FGRSADAIYLLYTGTTGTFPKGYVMRHHEDIYRVLFGGTDFATGEHYKDEYDLAKAAA ANPMIRYPIPMIHGATQSATWMALESQOTVLAPEFNADDEVWRTIHKHKNLLFTT GDAMARPLVDALVKGNNDYDLSLFLASTAALFSPSIKEKLELLELPNRVITDSIGSSE TGFPGTSVAAGQAHGCGPRVRIHRTVYLLDDDNVKNPKSGMGRVIAKKNIPGVSY KDKERTAEFTRTINGVRYAIPGDAQVEEDGTVMILGRGSVINSGEKVPYEEVEAAA LKHPDVFDAALVGVPPDPRIYGOQVAAYVQARPGCRPSLAELDSFYVSEIAGYKVPKRS LWFVEVYKRSAPGKPDYRWAKEQTEARADDVHAGHVTSG"	BASE COUNT	2292 a	5217 c	6167 g	1634 t	1 others
ORIGIN													
Alignment Scores:													
Pred. No.:	11.5	Length:	15311										
Score:	168.00	Matches:	96										
Percent Similarity:	37.67%	Conservative:	43										
Best Local Similarity:	26.02%	Mismatches:	132										

Query Match:	6.97%	Indels:	98
DB:	1	Gaps:	14
US-09-825-414-7 (1-486) x AE007164 (1-15311)			
QY 145	PROAlaProLeuLysProLeuThrProLeuIleSer--GlyAlaLeuSerGlyAlaMet	164	
DB 6720	 CCGGCGCG--CCGCTGCCCGCTTGAGTGAAGAGGCTGCCGCGCGCTCCG	6667	
QY 164	spGlnValGlyThrLysMetMetAspArgAlaThrGlyAspLeuHisThrLeuSerAla	184	
DB 6666	 GTGCGCGCGGACCGCCTTGCGCGCTTGCCAGCGGT-----	6629	
QY 184	erProAspArgLeuHisAspAlaMetAlaAlaSerValLysArgHisSerProSerLeu	204	
DB 6628	-----GCCGTTAGCGCGCCACCGCGCGCCAGCGCGCCAGCGCGCGC	6593	
QY 204	laArgGlnValLeuAspThrGlyValAlaValGlnThrTyrSerAlaArgAsnAlaVal	224	
DB 6592	-----CGCCCGACCGGACACCC-----	6575	
QY 224	rgThrValLeuAlaProAlaLeuAlaSerArgProAlaValGlnGlyAlaValAspLeu	244	
DB 6574	-----GTCGCGCGCTTGACCTCTTGCCCGCGCTTGCGCGCGGACTGTGTG	6526	
QY 244	lyValSerMetAlaGlyGlyLeuAlaAlaAsnAlaGlyPheGlyAsnArg-----	260	
DB 6525	 GGCGCGCT--GCCCGCGCGCGCGCGACCGCGGACTTCCGCGCGTGGCG	6469	
QY 261	--LeuLeuSerValGlnSerArgAspHisGlnArgGlyGlyAlaLeuValLeuGlyLeu	280	
DB 6468	 CCCTTGCGCGCGCGCGCGACCGCATACCTCGCGCGCGCGCGCCACCACT-----	6419	
QY 280	ysAspLysGluProLysAlaGlnLeuSerGluGluAsnAspTyrLeuGluAlaTyrLys	300	
DB 6419	-----	6419	
QY 300	laIleLysSerAlaSerTyrSerGlyAlaAlaLeuAsnAlaGlyLysArgMetAlaGlyL	320	
DB 6418	::::: -----GCCGCTCTGCGCGCGGTGCCGCTTGCGCGCGCGCGCGCGGTGCCGCC	6367	
QY 320	euProLeuAspMetAlaThrAspAlaMetGly--AlaValArgSerLeuValSerAla	339	
DB 6366	::: CCAACGTTGCTGCCAGCGGTGGCGCGCGCGCGCGCGCGCTTGCGCGCGCGCG	6307	
QY 339	erSerLeuThrGlnAsnGlyLeuAlaLeuAlaGly--GlyPheAlaGlyValGlyLysL	358	
DB 6306	::::: GTACCGTTGGCGCGCGACGCTACCGCGTGTGCTGCCCGCGGACACCGGCGCGT	6247	
QY 358	euGlnGluMetAlaThrLysAsnIleThrAspProAlaThrLysAlaAlaValSerGlnL	378	
DB 6246	 CCGCCCTTCCGCCAGT-----CCACCGTTTCCGCGCGCTATTTCGCTA	6202	
QY 378	euThrAsnLeuAlaGlySerAlaAlaValPheAlaGlyTyrThrThrAlaAlaLeuThrT	398	
DB 6201	 CCACCGCGCTCTCCGCCAGCGCGCGCGATGGCGCGCTTGCGCGCGCTGACCT	6142	
QY 398	hrAspProAlaValLysLysAlaGlySerPheIleGlnAspThrValLysSerThr----	416	
DB 6141	::: CCGGTCCCGCGCGCGCGCGCTGATGCGCTACCGCGACGACCGCGGTTCACCGTTG	6082	
QY 417	-----AlaSerSerThrThrGlyTyrVal	425	
DB 6081	CGCGCGCGCGCGCGGTTCGCCCATGCCCGCGCTGCCAGCGCGCGGTGCGCGGAGATG	6022	
QY 425	laAspGlnThrValLysLeuAlaLysThrValLysAspMetGlyGlyAlaIle----	443	
DB 6021	 CCGACGCGACACCGCGCGCATCTCGCGCGTCCGCCCATTCGCACGATACCGTTGGCG	5962	
QY 444	-----ThrHisThrGlyAlaSerLeuArgAsnThrValAsnAsnLeuArgGlnArgPro	462	
DB 5961	::: CCCGAGTGGCACCGCGCGCGC-----CGCGCGCGCGCTCTCCGCCCTTTCGCGCGCTGCCG	5905	

Oy	462	laarglualalaspillegluglully-----GlyThralalAserp	476
		::	
Db	5904	CCCCGGCCCCGCTGTGCGGGCGGTGCTGCTGACGCTGCTCACCGTCCGCGCT	5845
Oy	476	rosergluilepophearpro	483
		::	
Db	5844	TTGCCGCGCATTGCGCGCGCCCC	5822
RESULT 5			
MTV023/c			
LOCUS		47852 bp	DNA
DEFINITION			linear
ACCESSION			BCF 03-AUG-2001
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
COMMENT			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
COMMENT			
FEATURES			
source			
source			
source			

```

misc_feature
/ab_xref="taxon:833332"
/clone="Y3C7"
<1. .47852
/note="fragment designated v023. Does not represent a
physical clone"
complement(1. .1695)
/gene="Rv3494c"
complement(1. .1695)
/gene="Rv3494c"
/note="Rv3494c, (MTV023.01c), len: 564. Unknown Pro-rich
protein similar to several Mycobacterium tuberculosis
proteins e.g. MTC128.14 (515 aa), MTC19H5.28c (516 aa) and
(MTV051.09). Has hydrophobic stretch, possibly signal
peptide at N-terminus. FASTA scores: 297050|MTC128_14
(515 aa) opt: 1006 z-score: 582.8 E(): 5.9e-25;
identity in 559 aa overlap: 297182|MTC19H5_28 (516 aa)
opt: 979 z-score: 567.7 E(): 4.1e-24; 33.5% identity in
555 aa overlap. Tbpase score is 0.897"
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv3494c"
/protein_id="CA17731.1"
/db_xref="GI:2924431"
/ab_xref="SPTREMBL:O53539"
/translation="MIDRLAKIQLSIFAVITVITLSMAIFYRLRPATFGIGTVGSA
DEVAGGGLKMANVTRGVAAGRVESVGLNPNGVTAHMRLNSGTAIPSNVTATVRVS
AIGEQYIDLVPENPSSTKLRNGFRIRQONTIRIGODVADLROAETLGLSGDTRLRE
LHFAFIATNGAGPELARIIESARLVDANANYQVSQIIDQAGPFLQAIRAGDI
KSLADGLARFTWQLRADPRLRDLTADAPDAIDEANTAFSGIRPSFALAASLANGR
VGVIYHKSIEQLLVFPALFAAIITTSAGVPODEGAKLDFKIDLHDPCCMTGFLPP
LVRSAPADESVREIPDMYCKTAQNDPSTVRCARNYPCQEFPGKRAPTVQLCRDPGRYV
PRTNPMWRGPRPIPYGTEVTGDRNILLPNNKFPYIPGADDPGVPIVGGPPOVAGPG
PAPHQAPQAPAPPNDNGPPPPFTSMPPGYPPEPQVPYPATIPPPPEGTGPPPGP
APGPQASGPAYTIYDQSGAFADPAGGTGIFAPGMTGASSAENWVDMRDPRL"
complement(1703. .1707)
/note="possible RBS for Rv3494c"
complement(1706. .2860)
/gene="lprN"
complement(1706. .2860)
/gene="lprN"
complement(1706. .2860)
/note="Rv3495c, (MTV023.02c), len: 384. lprN, similar to
Mycobacterium tuberculosis proteins MTC128.13 (390 aa) and
MTCY19H5.29 (402 aa) and (MTV051.08). Probably
lipoprotein, contains possible signal sequence and
appropriately positioned PS00013 prokaryotic membrane
lipoprotein lipid attachment site. FASTA scores:
gpi297050|MTC128_13 (390 aa) opt: 653 z-score: 762.1 E():
0; 33.6% identity in 363 aa overlap; and
297182|MTCY19H5_29 (402 aa) opt: 572 z-score: 667.9 E():
1.1e-29; 31.8% identity in 362 aa overlap. Tbpase
score is 0.897"
/codon_start=1
/transl_table=11
/product="lprN"
/protein_id="CA17732.1"
/db_xref="GI:2924432"
/db_xref="SPTREMBL:O53540"
/translation="MNRIMLRALIIITASSALLACQFGGLNSLPLPGTAGHGEGATSV
TVMADVATLIPONSVMVVDVTGVSAGIVAVQRPDGSFYAAVKLDLKNVLLPANAV
AKVSQTSLLGSLHVELAPPTDRPPTGRLVDSGRITTEANTDRFPTTEVFSALGVVVK
GNVGALEEIIIDETHQAVAGROAQFVNLPRLAETLAGLNROVHDIIDLGLNRVSAI
LARDKDNLGRALDTLPDAVAVRLNQNRDHYDAFAALKRLTMVTSHYLAETKVDGEDL
KDIYSIVKALNDRKDFVTSIQLLTTFEPNFGIKQAVRGDYLVNFTEDTLRLRIGE
TFTTAYFDPNMAHMDETLNPDPFLIGELANLSGQAADPFKIPPGTASGO"
complement(2798. .2830)
/gene="lprN"
/note="PS00013 Prokaryotic membrane lipoprotein lipid
attachment site"
complement(2857. .4212)
/gene="Rv3496c"
complement(2857. .4212)
/gene="Rv3496c"
/note="Rv3496c, (MTV023.03c), len: 451. Unknown but

```

similar to Mycobacterium tuberculosis proteins MTC128_12 (530 aa), MTCY19H5_30c (508 aa) (MTV051_07). Hydrophobic region at N-terminus. FASTA scores: Z97050|MTC128_12 (530 aa) Opt:838 z-score: 711.0 E(): 4.3e-32; 35.1% identity in 473 aa overlap; and Z97182|MTCY19H5_30 (508 aa) Opt: 821 z-score:697.0 E(): 2.6e-31; 35.1% identity in 453 aa overlap.TBparse score is 0.891"

```

/codon_start=1
/translation_table=11
/product="hypothetical protein RV3496c"
/protein_id="CAAL7733.1"
/db_xref="GI:2924433"
/db_xref="SPTREMBL:O53541"
/translation="MGRVAMLTGSRGLRYATVIALVAALVGGVYVLLSFGNKRTIIVG
YFSAVGLPYGDQVRVLGPVGEIDMEPRSSDVKITMSVSKDVKPYDVQAVIMSPNN
LVAARFQLTPLYTGGAVLPDNGRIDRTPAVPEWDEVEKEGLTRLAIDSPAGELO
GPLGALINQADTLIDNGDSLHNALRELAQVAGRLGDSRGDIFGTYNLQVLDALSEE
SDEQIVQFAGHAVASQVILADSSANDPOTLGLTNOALSDIRGLRENNSTLIETVNL
NDFQTLTLDQSENIQVQLHVAGPGITNEYNIYDPAQGTNLGLSLIPNPNFVQFICGG
SFDPAGPSAPADYYRRAEICRERLGLVRLTLVNPPIMFHPLNTITAKGQIITYDP
ATEAKSETPVELTWPVAGGGA PVGNPADLQSLVLPAPGPAPAPAPGAGPGEHGGG
G"

```

RBS	complement(2867. .2871) /gene="Rv3496c" /note="possible RBS for Rv3495c"
gene	complement(4209. .5282) /gene="Rv3497c"
CDS	complement(4209. .5282)

```

/note="RV3497C, (MTV023.04c), len: 357. Unknown but
similar to Mycobacterium tuberculosis proteins MTCY19H5.31
(481 aa), MTCI28.11, (515 aa) and MTV051.06. Hydrophobic
region atN-terminus. FASFA scores: Z97182|MTCY19H5.31
(481 aa) opt:611 z-score: 667.7 E(): 1.1e-29; 32.5%
identity in 332aa overlap; and Z97050|MTCI28.11 (515 aa)
opt: 587 z-score:641.2 E(): 3.3e-28; 30.1% identity in
335 aa overlap. Tbpase score is 0.889"
/codon_start=1

```

gene
CDS

```

/transl_table=11
/product="hypothetical protein Rv3497c"
/protein_id="CA17734.1"
/db_xref="GI:2924434"
/db_xref="SPTREMBL:O53542"
/translation="MLNRPSSKHERDPLRTGIFGLVLIVICVLLIAFGYSGLPFWPQG
KTDAYFTDAGGITPGNSVSYGLKGVASAVSLAGSAKVTFSVDRSIYGDQSLAA
IRDTITLIGERSIAVSPAGSGKSTIPLSRITTPYTLNGVLQDLGRNANDLNRPFQEA
LNVEFGALHDATPOVRGADVGLTSLRALNRDEALQGLLAHAKSVTSLSRAEOVN
KLVEGQNLFAALDARRAALSALISGIDVAAQISGEVADNRKEFGPALSKNLVLAN
LNERDYITEALKRLPTVATTLGVEVSGPGFENVNYSVLPGLVATVDFVEQPGKL
PDSLADYLRGFIQERWIIIRKPS"
complement(5272..6324)
/gene="Rv3498c"
complement(5272..6324)

```

```

/note="Rv3498c, (MTV023.05c), len: 350. Unknown but
similar to Mycobacterium tuberculosis proteins MTC128.10
(346 aa), MTCY19H5.32c (346 aa) and MTV051.05. Hydrophobic
region at N-terminus. FASTA scores: 297050|MTC128.10 (346
aa) opt: 755 z-score: 883.6 E(): 0; 36.8% identity in
340 aa overlap; and 297182|MTCY19H5.32 Mycobacterium
tuberculosiscosmi (275 aa) opt: 699z-score: 820.1 E(): 0;
47.0% identity in 249 aa overlap. Tbpase score is 0.878"
/codon_start=1
/translate=11

```

```

Alignment Scores:
Pred. No.:      81.9
Score:          164.00
Percent Similarity: 37.40%
Best Local Similarity: 25.75%
Query Match:    6.80%
DB:             1

```

Length:	47852
Matches:	95
Conservative:	43
Mismatches:	133
Indels:	98
Gaps:	14

US-09-825-414-7 (1-486) x MTV023 (1-47852)

QY	145	ProIaPrProLeuLysProLeuThrProLeuIleSer--GlyAlaLeuSerGlyAlaMetA	164
Db	31455	CCGGCCGGCCG-----CCGGCTGGCCGGCTGTGGTAGCTGAGGAGGCTGGCCGGCCGGCCG	31402
QY	164	spgInValGlyThrLysMetMetAspArgAlaThrGlyAspLeuHisTyrLeuSerAla	184
Db	31401	GTGGCCCGCCGACCGCCTTGGCCGCGTGTGCCAGCGGT-----	31364
QY	184	erProAspArgLeuHisAspAlaMetAlaAlaSerValLysArgHisSerProSerLeuA	204
Db	31363	-----GCGGTAGCGCCGCCACCGCGCGCCGACGCCCGCCG	31328
QY	204	laArgGlnValLeuAspThrGlyValAlaValGlnThrTyrSerAlaArgAsnAlaValA	224
Db	31327	-----CGCCCCACCGGCACACC-----	31310
QY	224	rgThrValLeuAlaProAlaLeuAlaSerArgProAlaValGlnGlyAlaValAspLeuG	244
Db	31309	-----GCTGGCCGCTTGACCTCTCGTGGCCGCGGTGGCCGCGGAGCTGGTGTGG	31261
QY	244	lyValSerMetAlaGlyGlyLeuAlaAlaAsnAlaGlyPheGlyAsnArg-----	260
Db	31260	GGCCCGCT--GCCGCCGGCGCGCCGCCACCGCCGCACTTCCCGCTCCGCGCGCTGGCCG	31204
QY	261	--LeuLeuSerValGlnSerArgAspHisGlnArgGlyGlyAlaLeuValLeuGlyLeuL	280
Db	31203	CCCTTGGCCGCGCCGACCGCACCAGATACCGTCCGCCGCGCGCCACCACT-----	31154
QY	280	ysAspLysGluProLysAlaGlnLeuSerGluGluAsnAspTrpLeuGluAlaTyrLysA	300
Db	31154	-----	31154
QY	300	laIleLysSerAlaSerTyrSerGlyAlaAlaLeuAsnAlaGlyLysArgMetAlaGlyL	320
Db	31153	-----GCCGCTCTGCCGCGGTGGCCGCGCTTGGCCGCGCGGAGCCGCTGGCCGCC	31102
QY	320	euProLeuAspMetAlaThrAspAlaMetGly--AlaValArgSerLeuValSerAla	339
Db	31101	CCACCGTGTCTGCCAGCCGTGGCGGCCCGCCGACGCCGCGTGGCCGCGCTGGCCGCCG	31042
QY	339	erSerLeuThrGlnAsnGlyLeuAlaLeuAlaGly---GlyPheAlaGlyValGlyLysL	358
Db	31041	GTACCGTGTGGCGCCGACGAGTACCGCGCTGTGTCGCCGCGGACACCAACCGCACCGGC	30986
QY	358	euGlnGluMetAlaThrLysAsnIleThrAspProAlaThrLysAlaAlaValSerGlnL	378
Db	30985	-----GTTTCCGCGCTTTCGCCCAATCCCAACCGTTTCCGCGCGCTATTTCCGCTA	30937
QY	378	euThrAsnLeuAlaGlySerAlaAlaValPheAlaGlyTyrThrThrAlaAlaLeuThrT	398
Db	30936	CCACCGCGCTCTCCGCCACGCGCGCGGATGGCGCGGTGGCCGCGGTGGCCGCGGTACT	30877
QY	398	hrAspProAlaValLysLysAlaGluSerPheIleGlnAspThrValLysSerThr---	416
Db	30876	CCGGTCCCGCGCGGACGCGCGTGCATGCCGCTACCGCCAGCACCGCGGTTCCACCGTTG	30817
QY	417	-----AlaSerSerThrThrGlyTyrValA	425
Db	30816	CCGCGCGCGCGCGCGGTTCGCCCATGCCCGCGCGTCCGACGCGCGGTGGCGCGAGATG	30757
QY	425	laAspGlnThrValLysLeuAlaLysThrValLysAspMetGlyGlyGluAlaIle----	443
Db	30756	CCGACGGCACACCGCGCCATCTCCGCGCTCGCGCCCATTTGCCACGCGATACCGTTGGCG	30697
QY	444	-----ThrHisThrGlyAlaSerLeuArgAsnThrValAsnAsnLeuArgGlnArgProA	462
Db	30696	CCCGGAGTGGCACCGCGCGCGCC--CGCGCGCGCGTCTCCGCCCTTTCGCGCGCGTGGCG	30640
QY	462	laArgGluAlaAspIleGluGluGly-----GlyThrAlaAlaSerP	476
Db	30639	CCCGCGCGCGCTGTGGCGCGCGGTGCTGTGTGACCGCTGTCCACCGTGGCGCGCT	30580

Qy 476 rosergluilePropheargPro 483
Db 30579 TTGCCGCCATTGCCGCCGCCGCC 30557
RESULT 6
AE005780/c 14131 bp DNA linear BCT 12-JUN-2002
LOCUS Caulobacter crescentus CB15 section 106 of 359 of the complete genome.
DEFINITION AE005780 AE005673
ACCESSION AE005780 AE005673
VERSION AE005780.1 GI:13422307
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
1 (bases 1 to 14131)
Nierman,W.C., Feldblyum,T.V., Laub,M.T., Paulsen,I.T., Nelson,K.E., Eisen,J., Heidelberg,J.F., Alley,M.R., Ohta,N., Maddock,J.R., Potocka,I., Nelson,W.C., Newton,A., Stephens,C., Phadke,N.D., Ely,B., DeBoy,R.T., Dodson,R.J., Durkin,A.S., Gwin,M.L., Haft,D.H., Kolonay,J.F., Smilt,J., Craven,M.B., Khouri,H., Shetty,J., Berry,K., Utterback,T., Tran,K., Wolf,A., Yamathewan,J., Ermolaeva,M., White,O., Salzberg,S.L., Venter,J.C., Shapiro,L. and Fraser,C.M.
Complete genome sequence of Caulobacter crescentus Proc. Natl. Acad. Sci. U.S.A. 98 (7), 4136-4141 (2001)
21173698
PUBMED 11259647
REFERENCE
AUTHORS
2 (bases 1 to 14131)
Nierman,W.C., Feldblyum,T.V., Paulsen,I.T., Nelson,K.E., Eisen,J., Heidelberg,J.F., Alley,M.R.K., Ohta,N., Maddock,J.R., Potocka,I., Nelson,W.C., Newton,A., Stephens,C., Phadke,N.D., Ely,B., Laub,M.T., DeBoy,R.T., Dodson,R.J., Durkin,A.S., Gwin,M.L., Haft,D.H., Kolonay,J.F., Smilt,J., Craven,M., Khouri,H., Shetty,J., Berry,K., Utterback,T., Tran,K., Wolf,A., Yamathewan,J., Ermolaeva,M., White,O., Salzberg,S.L., Shapiro,L., Venter,J.C. and Fraser,C.M.
Direct Submission
Submitted (31-JAN-2001) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA
FEATURES
source
1. 14131
/organism="Caulobacter crescentus CB15"
/strain="CB15"
/db_xref="taxon:190650"
complement(84. 1373)
/gene="CC1016"
complement(84. 1373)
/gene="CC1016"
/note="similar to GP:6064113; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="lpse protein"
/protein_id="AAK23000.1"
/db_xref="GI:13422308"
/translation="MTSRLLETWRRLPTPIRSAHVAVAGAPRAALEALDKALAEHRHR SAERTALARARRAGPRGLSPPLPTVIGFHSAVHGLEGARMLARGFGDMGLGVRAL DLSAVGFAAEIAPAYSSPDPEGVITISHINPELLRWARETEGLEGRHIGYWA WELEEVPSDWLPAFDFVDEVETPSAFAADAIRVAPRGVKTVPYPLYNRPQADR QRFGLQDDRVVLMAFDLRSTQQRKNPDALRAFRDATVKATRPATLVCKVVGADLYP ETFQALAAEVADDPISIRLTBDNSAQDMAALTASSDIVLSHRSEGYGLLAEAIWLG KPTLATGWSNVFEFMDPASSQPEVDYRLVPVEGDVIYRAGRWADADVGDAAEKLARMI SDDAARNITLAAATARNGHVSFNDAVAVMTSARLPLT"
1398. .3062
/gene="CC1017"
1398. .3062
/gene="CC1017"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11

/product="conserved hypothetical protein"
/protein_id="AAK23001.1"
/db_xref="GI:13422309"
/translation="MGDRIVAOAEPPTPAKLRGASGDIVMRKVCMQWRPTARGEV STEWSAGYVTDVNTYTFGYGELNPLRCRLPLTVGRHAPAKIENACELGFGQLSVSIH AAQPGINWYGTDFNPQAAFAEMVRLSGAEAKLYDEAFAEFCNRKDLPDFDTGLH GIMTWISDONRHLVDFTIRKLRPGVLYISYNTLPWSSSAPIRHLMKRHADVWGP GGQVIGQTDALNFIIDQFLALDPLYAKVNPGAVERLKQYKAODRKLYLAHEYMRDWHP MYFADVENMVGDAKVGFCASAHVLDNLDNMTPOQTEFMRDVSDSLRETRDFCMN QQFRDYYWVGVRALIGHQVAAAREERVMTVAQADLPTRYRGVIGCATLNDVYQP LFEILDHKPHSVLELETHLASKSLHFGQLSALTILLGLGVIYTTAAAOVSKAKAR TLALNKTIARRSLSVGDIAYFASPITGGGTTASRFAQLFWLSRONGGKTAEDWASFAY QVLOGGQAIIKDGATLQGEDNLTLELRAQATHWGAHVLPWNA LGV"
complement(3751. .4500)
/gene="CC1018"
complement(3751. .4500)
/gene="CC1018"
complement(3751. .4500)
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK23002.1"
/db_xref="GI:13422310"
/translation="MPLDILAVTVFACCWLLYEPMLKRIGESGAVLNTDMTIVIRRW MQEMAVREIALLDGQLGHAINSAFASNNLILIAAAGVLEGGDNALKSVEGLAVL AKTTPLMFQIKGLVLVALRGLDFTWISIRQMNCLAAIGATPMWAPNAVLAETAEAGITNPLALSAFNAGVRAYFEALAAVWLGLPLFLTATIGAMLLMLLRORSSASVA VRKVRQLERQPVHGPHLSTSTSPEDRI"
complement(4542. .5975)
/gene="CC1019"
complement(4542. .5975)
/gene="CC1019"
complement(4542. .5975)
/note="identified by match to protein family HMM"
/codon_start=1
/transl_table=11
/product="major facilitator family transporter"
/protein_id="AAK23003.1"
/db_xref="GI:13422311"
/translation="MTATTTEATRLTPVARARAILGSGAGNIVEMWDFAYAAFTLYPA PAFFPKDQTVQLQAAAVFPLGVARPIGAWIMGLYADHAGRAALSVSVLMCAGA FIATIPDEQITIGLWAPAILLPARYQLGSLVGEGYASAPVSEMAKGARGFWSFHYVTIAGQLLALGVLIILQRTIGEEALKEMAWRPFTIGALLAVVFWIRRGLEESVS FKRPQOSETLSRQIISAGAFSLTJASGVGLMGSPILKAGQYLGAAFLVLFVSLI VPLIRHPRESLIMGLTAGSLFTFYVYTYTMQKFLVNTAGFSKAQASEISALSIAP MLMQPLAGWLSDRFGRKPMILITAFGGVLTWIPMTAISQNSVSAQALILVGAIFQ SCYTAISAVVKAEMFPAEIRALGVALPYALANVLFGTAENVALAFKHENLESTFYVY VAAVMTIGLICSITLKDTGRHSLIHED"
6053. .6442
/gene="CC1020"
6053. .6442
/gene="CC1020"
complement(6408. .6800)
/gene="CC1021"
complement(6408. .6800)
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK23004.1"
/db_xref="GI:13422312"
/translation="MRASLSRQNDSVLTREALEGEIRRTPGQKAGLEPAHQNDQDD HEDQAKAAAGEIAPRPVPPGGORADOGODQKNENNENRAEHGPKVPVEPASATVAREVM GSEFAAMLVGSEPKGSLKRVRLQPCGS"
complement(6408. .6800)
/gene="CC1021"
complement(6408. .6800)
/gene="CC1021"
complement(6408. .6800)
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAK23005.1"
/db_xref="GI:13422313"
/translation="MDAMSEPPDPSVLTGLKRGLRHRCNCGDGRLYRRYLKVSRYC EACGHELGLYPADDPAYFTIFVGHIIIIAPMLFFPWIMWASFWLVVPTLIPRLALT LTIIPRVKGAVIGTLMWAIKIRKADAPS"

```

gene
6863. .7117
/gene="CC1022"
CDS
6863. .7117
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK23006.1"
/db_xref="GI:13422314"
/translation="MLGVARGADSLAGRHIAAHALDRVARRNQRRGGHGEDDELA
HIGNSTRSRGWRTLALRTELDQEPGAFFALFSELGSP"
complement(7092. .7508)
/gene="CC1023"
CDS
complement(7092. .7508)
/note="identified by match to protein family HMM"
/gene="CC1023"
/codon_start=1
/transl_table=11
/product="beta/gamma crystallin family protein"
/protein_id="AAK23007.1"
/db_xref="GI:13422315"
/translation="MRKTFPLAGGALTTLATQALAQDRDWRDRHSEQRDDRYDRR
DDRDDRDRDWGGGIVLEFHRDFYGDARPIRGDMPDLARIGFNDRVSSMRISRGAQW
FCEHAYFOGRCWVYDYGVPENRQNDRFSSVRVR"
complement(7519. .8073)
/gene="CC1024"
CDS
complement(7519. .8073)
/note="identified by match to protein family HMM"
/gene="CC1024"
/codon_start=1
/transl_table=11
/product="beta/gamma crystallin family protein"
/protein_id="AAK23008.1"
/db_xref="GI:13422316"
/translation="MAGSAAQAQPYWNGRYDDPPRGSYTRSCRETTAFNGQVWARCQT
DRGWEWSSARVREGCAQGIENRDRLQACAGFVGPGGGGGGGRPPYGGGGGY
RGDILMFEDTGYGGRVYEVGDTPLDLIARFNDRASSIRIVRGWICEHANYQGRCS
RLDADQFVLPREWDAISIRVR"
complement(8216. .10285)
/gene="CC1025"
CDS
complement(8216. .10285)
/note="identified by match to PFAM protein family HMM
PF01804"
/codon_start=1
/transl_table=11
/product="penicillin amidase family protein"
/protein_id="AAK23009.1"
/db_xref="GI:13422317"
/translation="MALVLAALALIAIDRASLPQAPSRQALIDKAKAYDVIRRDEW
GVPHILGRTDADAFFGLGQAOSDEDDTLQDVVLATRGVLAARKGAGAAPTDYYVSL
EWPTVDKHYAELPAELRLAIMEAYADGVSLYVAKHPEKAKPGLPVTGKDVAAAGFIFK
QPEFYGLDGEELRRTAPATKGPSPPKGSGNLATAPSRSDGAPTRLVNSHQPTGPVAV
YEAVVSGGQWVHAGGFEPGSEFMLGHNATLGWANTVSKPLDYVYRLTINPANKNO
YRIDGQWKDFDKRYVTLRYKLLGPIVLVRKAVLRSAHGPVLETDHGVEAIRYAGMGE
WRQPLQYWLNRARTEVEDMRAIATHAIPSLNVYADAGNIGFYVHNGQYPNRAKAD
WGVLPGDRSLDIWQGYLPEPNSPOLWNPKSGLVFNSNTPFEASEADNLKPADFPA
SMGLQANMTNRAWRALETYGADRALTDASFRAHKFDVAEFSRDSVMAVGEVLAVDPK
GDADLAEAQRILRAWDKRADRTNRGALAALMSOPILFANTNGDPAPAPIDSLRAAIK
TLKTHFGRLDPEWGEVNRIRRGTVLDLPIDGAADTFERSVWGKPKQDKGTTTADGGDTFVM

```

Alignment Scores:

Pred. No.:	20.7	Length:	14131
Score:	162.50	Matches:	136
Percent Similarity:	35.32%	Conservative:	77
Best Local Similarity:	22.55%	Mismatches:	217
Query Match:	6.74%	Indels:	174
DB:	1	Gaps:	29

US-09-825-414-7 (1-486) x AE005780 (1-14131)

QY 10 PROVALTHRALTHRASP-----SERPHEARGTHRALASERASPALASERLEUALA 27

Db	13542	CCCCCGCCGCTGTCGACCGATCCCTCCGACCTGAAGCCCGCAGCTGGCGGCTGGCG	13483
QY	28	SerSerSerValArgSerValSerSerAspGlnGlnArgGluIleAsnAlaIleAsp	47
Db	13482	---CGGCGCATCCAGTCCGGGGCCGACCGACGAGCCGCTGAGGCTGGCCACCCTGGCCGAC	13426
QY	48	TyrIleuThrAspHisValPheAlaAlaHisLys-----LeuProProAlaAsp	63
Db	13425	CGGCTGACGGCGGCTGGCTCGGCCGACGAGGAAGGCTGGGTCGTTCCGGGGGGCGGC	13366
QY	64	SerAlaAspGlyGlnAlaAlaValAspValHisAsnAlaGlnIle----ThrAlaLeu	81
Db	13365	GCGGCGGACGGCGATGAGCCCTATGACGCCGAGCGGGCGCGCTGTGATGACCGGCTTC	13306
QY	82	IleGluThrArgAlaSerArgLeuHisPheGluGlyGluThr-----	95
Db	13305	GTCGACTGGCGCGCCGACGCGGGCCAGGTCGACGACGAGACGGCTACATGTCCGCTCG	13246
QY	96	-----ProAlaThrIleAlaAspThrPhe---	103
Db	13245	ATCTACCTGATCGCCCTGGCCATGCTGTGTGAAGCCCGACACCGTGCCGGGCTGTTCAG	13186
QY	104	-----AlaLysAlaGluLysLeuAspArgLeuAlaThrThrSerGlyAlaLeuArg	121
Db	13185	CGGCGGCTCAGCGCCTTTCGCCGCGACTATCTGGCGGAGAGACGATCCGAGCGCTTGAG	13126
QY	122	AlaThrProPheAlaMetAlaSerLeuLeuGlnTyrMetGln-----	135
Db	13125	GCCGCGAGATCATCACCGAGCGCTGCTGCCGAGCAGAGAGTTGGTGGCCATGGACTTC	13066
QY	136	-----ProAla-----IleAsnLysGlyAspTrpLeuPro-----	145
Db	13065	ACGGCGGCGGAACCCCGCGCCCGCAAGCTGCCGCGCACCTGGCCGCCGCCGCGCA	13006
QY	146	-----AlaProLeuLysProLeuThr	152
Db	13005	CCGCCCTACCCCTCACGCAAGACCTGGCCAGAGAGGCCACGCCCTGGCCGCCGCGAG	12946
QY	153	ProLeuIleSerGlyAlaLeuSerGlyAlaMetAspGlnValGlyThrLysMetMetAsp	172
Db	12945	CCG-----	12943
QY	173	ArgAlaThrGlyAspLeuHisTyr-----LeuSerAla	183
Db	12942	---CGGCGCGGGCGGCTGCGGTATGGCGCGCGCCCTAGAGAGACGGCCCTCCCTCGCC	12886
QY	184	SerProAspArgLeuHisAspAlaMetAlaAlaSerValLysArgHisSerProSerLeu	203
Db	12885	TCAACCCGAACGCGCGCTTGAGCCGATA-----GAACCGCGCTC	12847
QY	204	AlaArgGlnValLeuAsp-----ThrGlyVal-AlaValGln---Th	216
Db	12845	GCCCGAGAGTGCATAGACGGCGCGGATGCGGAAGCGCTGGGGCTCTGGAGAGCTC	12787
QY	216	TyrSerAlaArgAsnAlaValArgThrValLeuAlaProAla-----	230
Db	12786	GTAATTCGGCTGGGACAGAGCCGAAGGGGCGACCTCGCCGGCAGGGTGCTTGACACAC	12727
QY	231	-----LeuAlaSerArgProAlaValGlnGlyAlaValAspLeuGlyValse	246
Db	12726	GATCAGGCGCGGATCTCGTCGAGGCCGAAGCTGATCACGTCCGACCCCTGGGGTCGCC	12667
QY	246	rMetAlaGlyGlyLeuAlaAlaAsnAlaGlyPheGlyAsnArgLeuLeuSerValGlnse	266
Db	12666	GCCCGCGCGGGCGAATGGCGCGGCCACGTCCGGCGCA-----CCGCGATC	12619
QY	266	rArgAspHisGlnArgGlyGlyAlaLeuValLeuGly-----LeuLysAspLys	282
Db	12618	GGAGAGCGCGCGCTCGTGGCGGACGATCTGGCGATCCCGAGGCCCTTGAGGAAGCG	12559
QY	282	sgluProLysAlaGlnLeuSerGluGluAsnAspTrpLeuGluAlaTyrLysAlaIle--	301

Db	12558	GTCTGACGCTCGCGGCTTCCCATGTCTGCAACCCGCTGGATGACGCGCTGGGT	12499
QY	302	-----LysSerAlaSerTyrSerGlyAlaAlaLeuAsnAlaGlyLys	315
Db	12498	CATCGGCATCGGCTTGGGGGGGGCGGCTCGATCAGTGGGACGGCGCTGGGGCTGTGAA	12439
QY	315	SArgMetaGlyLeuProLeuAspMetaIaThrAspAlaMetGlyAlaValArgSerLe	335
Db	12438	CAGGGCGGGGG-----CGACGGCGGGGCCAGCAGGGCGGGCTT	12397
QY	335	uValSerAlaSerSerLeuThrGlnAsnGlyLeuAlaLeuAla-----GlyG1	351
Db	12396	GCGGCGCAGCTGCTCTCCGACCGCGCTCGAGCAGGCGCTTGCATAGCCACGCGGGGG	12337
QY	351	YPheAla---GlyValGlyLysLeuGlnLumetaIaThrLysAsnIleThrAsp-----	368
Db	12336	CTGCTCTTGAAGATCGGCACGCCGATCAGCAGCAGCAGCCGAGATCTCGCCCTGCTT	12277
QY	369	-----ProIaThrLysAlaAlaVal-----Se	376
Db	12276	CCAGCTGATCGGGCGCTCGGCTTGCCGGTGACGAAAGCGGCGGTGCTGCGCGGC	12217
QY	376	rgLnuLeuThrAsnLeuAlaGlySerAlaAlaValPheAlaGlyTyrThrThrAlaAlaLe	396
Db	12216	GCGGTGACGGGCTTG-----CCGGCCAGGACCTTTCCAGCTGGGGCAGACATAGTCGCC	12163
QY	396	uThrThrAspProAlaValLysLysAlaGluSerPheIleGlnAspThrVal-----	413
Db	12162	CACCACCACTGCGAGTCCGCCCCCGTCCAATCGCGTCCGCCGATTCGCGCTCATGGG	12103
QY	414	-----LysSerThrAlaSerSerThrThrGlyTyrValAlaAspGlnThrValLys	430
Db	12102	ACGATGATGCGTTAGCGGGCGCGCGCGCCAGCGGT-----GAAGCTCAGCG	12055
QY	430	sLeuAlaLysThrValLysAspMetGlyGlyLysAlaIleThrHisThrGlyAlaSerLe	450
Db	12054	GCTTGCGCGGATGGCCATGTCTATGCGCGGGCGTCTCCGACGGGGGTCTCGTCTCTT	11995
QY	450	uArgAsnThrValAsnAsnLeuArgGlnArg-----ProIaArgGluAlaAspI1	467
Db	11994	GAGGCCCGCGCGCTCATGCGCGCGCGCAGGGGTGTCACAGGGGCTCGGC-gccGACCT	11936
QY	467	egLnuLys-----GlyGlyThrAlaAlaSerProSerGluIleProPheAr	482
Db	11935	TGAAGTCAAGATGCGGCTTGCGGGGATGTGTCCCGAGGTTGGAAAAAGCCCCC---CG	11879
QY	482	gPromet	484
Db	11878	CCCCGCTC11872	

RESULT 7	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS
SC4A7	42526 bp	DNA	linear	BCT 12-MAY-2002	
Streptomyces coelicolor	cosmid 4A7.				
AL133423	AL64582				
AL133423.2	GI:20520774				
acetyltransferase; acyl carrier protein; aldo/keto reductase; aldose 1-epimerase; aminotransferase; beta-ketoacyl synthase; beta-ketoacyl-acyl carrier protein synthase III; dehydratase; dehydrogenase; integral membrane protein; lipoprotein; malonyl CoA:acyl carrier protein malonyltransferase; marR-family transcriptional regulator; membrane protein; merr-family transcriptional regulator; oxidoreductase; putative aldose 1-epimerase; secreted protein; sugar transport permease; sugar-binding receptor; sugar-transport ATP binding protein; tetracenomycin C resistance and export protein; transcriptional regulator.					
SOURCE	Streptomyces coelicolor A3(2).				
ORGANISM	Streptomyces coelicolor A3(2)				
	Bacteria; Firmicutes; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.				
REFERENCE	1 (bases 1 to 42526)				
AUTHORS	Redenbach, M., Kleser, H.M., Denapaite, D., Eichner, A., Cullum, J.,				

TITLE	A set of ordered cosmids and a detailed genetic and physical map for the 8 Mb <i>Streptomyces coelicolor</i> A3(2) chromosome
JOURNAL	Mol. Microbiol. 21 (1), 77-96 (1996)
MEDLINE	97000351
PUBMED	8843436
REFERENCE	2 (bases 1 to 42526)
AUTHORS	Seeger, K.J. and Harris, D.
JOURNAL	Unpublished
REFERENCE	3 (bases 1 to 42526)
AUTHORS	Cerdeno, A.M., Parkhill, J., Barrell, B.G. and Rajandream, M.A.
TITLE	Direct Submission
JOURNAL	Submitted (09-DEC-1999) <i>Streptomyces coelicolor</i> sequencing project,

COMMENT On May 9, 2002 this sequence version replaced gl:6562773.

Notes:

Streptomyces coelicolor sequencing at The Sanger Centre is funded by the BBSRC and Beowulf Genomics. Details of *S. coelicolor* sequencing at the Sanger Centre are available on the World Wide Web.

(URL: http://www.sanger.ac.uk/Projects/S_coelicolor/) CDS are numbered using the following system eg SC7B7.01c. SC (S. coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary strand).

The more significant matches with motifs in the PROSITE database are also included but some of these may be fortuitous. The length in codons is given for each CDS.

usually the highest scoring match found by fasta -o is given for CDS which show significant similarity to other CDS in the database. The position of possible ribosome binding site sequences are given where these have been used to deduce the initiation codon. Gene prediction is based on positional base preference in codons using a specially developed Hidden Markov model (Krogh et al., Nucleic Acids Research, 22(22):4768-4778(1994)) and the FramePlot program of Bibb et al., Gene 30:157-66(1984) as implemented at

<http://www.nih.gov/jp/jun/cgi-bin/frameplot.pl>. CAUTION: We may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gtg, ttg or (att)) which is preceded by an upstream ribosome binding site sequence (optimally 5-13bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.

IMPORTANT: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions. Cosmid 4A7.

FEATURES

Source

/organism="Streptomyces coelicolor A3(2)"

/db_xref="taxon:100226"

626, 630

637. .1320

```
/note="SCC
```

637. .1320

```
/gene="SC4A7.02"
```

```

/note="SC4A7.02, probable transcriptional regulator, len:
227 aa; similar to SW:TCMR_STRGA (EMBL:M80674)

```

Streptomyces glaucescens tetracenomycin C tran

repressor TCMR, 226 aa; fasta scores: opt: 594 z-score:

727.4 E(): 0; 46.5% identity in 215 aa overlap. Contains Pfam match to entry PF00440 tetr, Bacterial regulatory

proteins, tetr family and a possible helix-turn-helix motif at residues 49..70 (+3.04 SD)"

```
/codon_start=1
/transl_table=11
```

```

/product="putative transcriptional regulator"
/protein_id="CAB62706.1"

```

misc_feature
/db_xref="GI:6562775"
/translation="MKVAGTSAPEETAPASRPGRLRELKKQRTDILLRSALLETTER
GVEETIVDDIAEADVSQRTFFRYEKEDAAFFVARIESHFYRAVLARPEEAPLD
ALRRLAESWSTIGEAVEQLVPLELHMRFYRVIESTPALLAHLRRATELEEIARV
AVREGLDVDTDPRPRIVAVEGAVMRTERIMWARDASLASLARDLTADYLDQVAPAL
TGNMRAEET"
733.858
/gene="SC4A7.02"
/note="Pfam match to entry PF00440 tetr, Bacterial
regulatory proteins, tetr family, score 64.10, E-value
4.5e-16"
1375.2595
/gene="SC4A7.03"
/note="SCO2375"
1375.2595
/gene="SC4A7.03"
/note="SC4A7.03, possible secreted protein, len: 406 aa;
similar to various hypothetical proteins, e.g. TR:Q9X8R0
(EMBL:AL049754) Streptomyces coelicolor hypothetical 33.9
kd protein, SCH10.21c, 329 aa; fasta scores: opt: 634
z-score: 714.7 E(): 2e-32; 39.3% identity in 346 aa
overlap. Contains possible N-terminal signal peptide
sequence"
/codon_start=1
/transl_table=11
/product="putative secreted protein"
/protein_id="CAB62707.1"
/db_xref="GI:6562776"
/db_xref="SPTREMBL:Q9RDQ9"
/translation="MTSEDTSPQNLNWRVLLALAVVEVMLATTGTALRDQRPATLE
ASVTAMEHGRIDGRLLPDARATPRILARFASLTAWQRTSLAHRYPLAVGNMGAPVQ
LRYLANRSALAEKARSVERETHDKRLSPAGOREAGRMHNYEALLAPGRHILAFDPAG
SGRAVEYFNLDRAEKRSVVPVGVTELLTFTORTDKRYAAPVGMAKSLYAERAASP
GTGTAVIAMADYTPSPGLMEATANRAAHGAVRLNALRLPGRSPVSLFCHSYGSV
VCGLAADMLPSRVTDIAVAGSPGMAENARLDT SARVWAMRDWDIODVPLYEVG
LHGADPVASAGFARVLSARDAQGHSGYFVPGTDSLRFNAGIGVGYRTVACAGEDDT
CRAGLSAATAAGRA"
2662.3549
/gene="SC4A7.04"
/note="SCO2376"
2662.3549
/gene="SC4A7.04"
/note="SC4A7.04, hypothetical protein, len: 295 aa;
similar to TR:Q9L067 (EMBL:AL163641) Streptomyces
coelicolor hypothetical protein SCCL05.18, 298 aa; fasta
scores: opt: 556 z-score: 591.0 E(): 2.8e-25; 37.415%
identity in 294 aa overlap"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein SC4A7.04"
/protein_id="CAB62708.1"
/db_xref="GI:6562777"
/db_xref="SPTREMBL:Q9RDQ8"
/translation="MGDVLAGFHAMFEESDVLIRYERGITRPKLFQALGERRVPLE
AIEGYTLPGRGTVLVREPRAGADPLIEAAGOLKEGDPYRLVLPADRETLAEY
ADELRALITESGRTERYLVAPPEAPLHFKAIDGKASFDGKAVSFRWSWTGASSAKWA
GDOSFPVADLIGVEMRSPMEFEGYRLPRDADASVAGAAGAPAVPOADODPAAYV
FGLGYPVHESLPFAAAVLAVERGASAPVIPVPAPRRDPADIAERIRHIGELHQ
GLVTDEEFSKKAELIAEL"
complement(3675.4691)
/gene="SC4A7.05c"
/note="SCO2377"
complement(3675.4691)
/gene="SC4A7.05c"
/note="SC4A7.05c, probable aldo/keto reductase, len: 338
aa; similar to TR:Q9X5G7 (EMBL:AF124928) Streptomyces
clavuligerus putative aldo/keto reductase family 2 enzyme
Cvml, 344 aa; fasta scores: opt: 916 z-score: 1035.7 E():
0; 47.7% identity in 331 aa overlap"
/codon_start=1
/transl_table=11
/product="putative aldo/keto reductase"

gene
complement(4688.5278)
/gene="SC4A7.06c"
/note="SCO2378"
complement(4688.5278)
/gene="SC4A7.06c"
/note="SC4A7.06c, possible merr-family transcriptional
regulator, len: 196 aa; similar to SW:TIPA_STRLI
(EMBL:S64314) Streptomyces lividans transcriptional
activator TipA, 253 aa; fasta scores: opt: 187 z-score:
225.2 E(): 3.7e-05; 32.7% identity in 110 aa overlap.
Contains Pfam match to entry PF00376 merr, Bacterial
regulatory proteins, merr family and a possible
helix-turn-helix motif at residues 72..93 (+3.19 SD)"
/codon_start=1
/transl_table=11
/product="putative merr-family transcriptional regulator"
/protein_id="CAB62710.1"
/db_xref="GI:6562779"
/db_xref="SPTREMBL:Q9RDQ6"
/translation="MSRGFSDDGFPPKALAWSALQHSVTVMQTAGTLVFNATS
NATGTAAGTDVCSAPPRIHPRPDGQDSYTISEVAFGLTAHLRWYERIGLMSHIDR
SHTGQRRYSNRDLWDLDFVGLRMTGMPVADWRYAEWVREGESTYPPDRRELESTRR
DVLTRIAELQDTVAVLDRIKISFYGDAGRAREGEGNR"
Alignment Scores:
Pred. No.: 110 Length: 42526
Score: 160.50 Matches: 125
Percent Similarity: 36.33% Conservative: 65
Best Local Similarity: 23.90% Mismatches: 212
Query Match: 6.65% Indels: 122
DB: 1 Gaps: 22
US-09-825-414-7 (1-486) x SC4A7 (1-42526)
QY 6 ArgValGlnGlnProProValThrAlaThrAspSerPheArgThrAlaSerAspAlaSer 25
||| ::| ||||| ||| ::| |||
Db 10392 CGGCGGAGACCGCGACACACGCGCCAGCAAGCGGCGGAGCGCGCCG 10451
QY 26 LeuAlaSerSerValArgSerValSerSerAspGlnArgGluIleAsnAlaIle 45
::| ::| ||| ::| ::| ::|
Db 10452 GACAAGGCCAAGCCCTGCGCGACGAGCCCTGGAGACGCCGAGCAGACGACGCCGCC 10511
QY 46 -----AlaAspTyrIleThrAspHisVal 53
Db 10512 CGGGCCAAAGCCGATGCCAAGGAGGCCCTATGCCCAAGGCCGTCCGCCGAGGACAATGCCG 10571
QY 54 PheAlaAlaHisIleuProProAlaAspSerAlaAspGlyGlnAlaAlaValAspVal 73
||| ||| ::| ::| ||||| |||
Db 10572 CAGGAGCGCCGAGAG-----GCGGCCAAGCAGCGCCGAGCGCG--GCCGCCGAGCAGCGCG 10622
QY 74 HisAsnAlaGlnIleThrAlaLeuIleGluThr--ArgAlaSerArgLeuHisPheGlu 92
||| ::| ||| ||| ::| ||| ::|
Db 10623 GAGACGGCGCGCGCGCGACGTCGCGAGCGCGGACAGGCGACGACGACGACGCGCGAT 10682
QY 93 GlyGluThrProAlaThrIleAlaAspThrPheAlaAlaGluAlaGluIleuAspArgLeu 112
::| ||||| ||| ::| |||
Db 10683 GCCGACGCGCGCGCACCTCGGCGAGCGCGCGGAGAGCGGTCCCGCGCGACTCCGAC 10742
QY 113 AlaThr-----ThrThrSerGlyAlaLeuArgAlaThrProPheAlaMet 127
||| ||| ||| ::| |||
Db 10743 GCCGCCAGCGCCGCAAGCTGACGCGCGGACGCGCGGTGAGGACCGCGCAGTCCGGCG 10802
QY 128 AlaSerLeuLeuGlnTyrMetGlnProAlaIleAsnIleuGlyAspTyrIleuProAlaPro 147

Db 10803 GCGGATGCCATCAAGCGCTCCAAGACCGCGCGGCGGCC----- 10844
QY 148 LeuLysProLeuThrProLeuIleSerGlyAlaLeuSerGlyAlaMetAspGlnValGly 167
Db 10845 -----CGTACGGCG 10853
QY 168 ThrLysMetMetAspArgAlaThrGlyAspLeuHisTyrLeuSerAlaSerProAspArg 187
Db 10854 GTGGAGTTGGCCGACGACGCC-----GAGCAGCAC-----GCCCGCGACGCC 10895
QY 188 LeuHisAspAlaMetAlaAlaSerValLysArgHisSerProSerLeuAlaArgGlnVal 207
Db 10896 AAGAAGGAGCGCGACCGCGCCCAAGCGCGAGGGGTCCACC--GCGCTCGCGGTGCCAAC 10952
QY 208 LeuAspThrGlyValAlaValGlnThr-----TyrSerAlaArgAsnAlaVal 223
Db 10953 GAGTCCACGGGCTACGCTACACACCACCGCGCAGCGCGGTGATGCCGCACTCGGCC 11012
QY 224 ArgThrValLeuAlaProAlaLeuAlaSerArgProAlaValGln--GlyAlaValAsp 242
Db 11013 GCGCAGGTAGCCGACCGCGCC-----AACGACGCCATCCAGCTCGGATCGCCGTAT 11063
QY 243 LeuGlyValSerMetAlaGlyGlyLeuAlaAlaAsnAlaGlyPheGlyAsnArgLeuLeu 262
Db 11064 GTCACACACCGACTCGCGCGCGCGGACTCGCGCTACTGACCGGACAGTCTCGAAGACGATC 11123
QY 263 SerValGlnSerArgAspHisGlnArgGlyGlyAlaLeuValLeuGlyLeuLysAspLys 282
Db 11124 GCCGAACAG-----CAGCAGGCCGTGCT----- 11147
QY 283 GluProLysAlaGlnLeuSerGluGluAsn----- 292
Db 11148 GAGGCGCAGCGCGCAGAACCGCGAGAGTGGCGCGCCCGCAGGACAGAGCGTGGCCAAACGCG 11207
QY 293 -----AspTyrLeuGluAlaTyrLys--AlaIleLysSerAlaSerTyrSerGly 308
Db 11208 GCGAGCGGTGACTCCAAAGCGCCCTACACACTGCGCGCGCAAGCCCTCGGCTACGCGCG 11267
QY 309 AlaAlaLeuAsnAlaGlyLysArgMetAlaGlyLeuProLeuAspMetAlaThrAspAla 328
Db 11268 GACGCGCGCACTCTGCCAAGAAAGCCTTGGGGTACTCGGACAGAAGCAGCCCTCGTACGCG 11327
QY 329 MetGlyAlaValArgSerLeuVal----- 336
Db 11328 ACCCAGGCGCCACAGTCACTCACCCGACCATCGCCTACGACACCCAGGCCACCAAGGAC 11387
QY 337 -----SerAlaSerSerLeuThrGlnAsnGlyLeuAlaLeuAlaGlyGly 351
Db 11388 GCGGACGCGCGCGACAGCG 11447
QY 352 PheAlaGlyValGlyLysLeuGlnGluMetAlaThrLysAsnIleThrAspProAlaThr 371
Db 11448 TCCGCCGATGCCGACCGCTGGACGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 11507
QY 372 LysAlaAlaValSerGlnLeuThrAsnLeuAlaGlySerAlaAlaValPheAlaGlyTyr 391
Db 11508 CAGCGCGCGGAGAC-----GCCCGGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 11546
QY 392 ThrThrAlaAlaLeuThrThrAspProAlaValLysLysAlaGluSerPheIleGlnAsp 411
Db 11547 GCCACCGAGCGCGCGCGCGCGGAGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 11606
QY 412 ThrValLysSerThrAlaSerSerThrThrGlyTyrValAlaAspGlnThrValLysLeu 431
Db 11607 GCTCAACAGCG-----GCCGAGCAGGCGCGGAGAGGAG 11639
QY 432 AlaLysThrValLysAspMetGlyGlyGluAlaIleThrHisThrGlyAlaSerLeuArg 451
Db 11640 GCAAAACGCGCAACAGATCGACAAGGAGAACGCGTGTGACACGACCGGAGCTCCCAT-CGG 11698
QY 452 AsnThrValAsnAsnLeuArgGlnArgProAlaArgGluAlaAspIleGluGlyGly 471

Db 11699 CGACGTGTT-----CTACGTGTGTCGACCACTCGAGAA--GATGCGCGAACCGGAGGT 11749
QY 472 ThrAlaAla 474
Db 11750 CGTGAAGCA 11758
RESULT 8
AE012164
LOCUS
DEFINITION
Xanthomonas campestris pv. campestris str. ATCC 33913, section 72 of 460 of the complete genome.
ACCESSION
AE012164 AE008922
VERSION
AE012164.1 GI:21111662
KEYWORDS
Xanthomonas campestris pv. campestris str. ATCC 33913.
SOURCE
Xanthomonas campestris pv. campestris str. ATCC 33913
ORGANISM
Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group; Xanthomonas.
REFERENCE
AUTHORS
1 (bases 1 to 10275)
da Silva,A.C.R., Ferro,J.A., Reinach,F.C., Farah,C.S., Furlan,L.R., Quaggio,R.B., Monteiro-Vitorello,C.B., Van Sluys,M.A., Almeida Jr.,N.F., Alves,L.M.C., do Amaral,A.M., Bertolini,M.C., Camargo,L.E.A., Camarotte,G., Cannavan,F., Cardozo,J., Chambergro,F., Ciapina,L.P., Ciccarelli,R.M.B., Coutinho,L.L., Cursino-Santos,J.R., El-Dorry,H., Faria,J.B., Ferreira,A.J.S., Ferreira,R.C.C., Ferro,M.I.T., Formighieri,E.F., Franco,M.C., Greggio,C.C., Gruber,A., Katsuyama,A.M., Kishi,L.T., Leite Jr.,R.P., Lemos,E.G.M., Lemos,M.V.F., Locali,E.C., Machado,M.A., Madeira,A.M.B.N., Martinez-Rossi,N.M., Martins,E.C., Melandis,J., Menck,C.F.M., Miyaki,C.Y., Moon,D.H., Moreira,L.M., Novo,M.T.M., Okura,V.K., Oliveira,M.C., Oliveira,V.R., Pereira Jr.,H.A., Rossi,A., Sena,J.A.D., Silva,C., de Souza,R.F., Spínola,L.A.F., Takita,M.A., Tamura,R.E., Teixeira,E.C., Tezza,R.I.D., Trindade dos Santos,M., Truffi,D., Tsai,S.M., White,F.F., Setubal,J.C. and Kitajima,J.P.
TITLE
Comparison of the genomes of two Xanthomonas pathogens with differing host specificities
JOURNAL
Nature 417 (6887), 459-463 (2002)
MEDLINE
22022145
PUBMED
12024217
REFERENCE
2 (bases 1 to 10275)
da Silva,A.C.R., Ferro,J.A., Reinach,F.C., Farah,C.S., Furlan,L.R., Quaggio,R.B., Monteiro-Vitorello,C.B., Van Sluys,M.A., Almeida Jr.,N.F., Alves,L.M.C., do Amaral,A.M., Bertolini,M.C., Camargo,L.E.A., Camarotte,G., Cannavan,F., Cardozo,J., Chambergro,F., Ciapina,L.P., Ciccarelli,R.M.B., Coutinho,L.L., Cursino-Santos,J.R., El-Dorry,H., Faria,J.B., Ferreira,A.J.S., Ferreira,R.C.C., Ferro,M.I.T., Formighieri,E.F., Franco,M.C., Greggio,C.C., Gruber,A., Katsuyama,A.M., Kishi,L.T., Leite Jr.,R.P., Lemos,E.G.M., Lemos,M.V.F., Locali,E.C., Machado,M.A., Madeira,A.M.B.N., Martinez-Rossi,N.M., Martins,E.C., Melandis,J., Menck,C.F.M., Miyaki,C.Y., Moon,D.H., Moreira,L.M., Novo,M.T.M., Okura,V.K., Oliveira,M.C., Oliveira,V.R., Pereira Jr.,H.A., Rossi,A., Sena,J.A.D., Silva,C., de Souza,R.F., Spínola,L.A.F., Takita,M.A., Tamura,R.E., Teixeira,E.C., Tezza,R.I.D., Trindade dos Santos,M., Truffi,D., Tsai,S.M., White,F.F., Setubal,J.C. and Kitajima,J.P.
TITLE
Direct Submision
JOURNAL
Submitted (28-NOV-2001) Departamento de Biolumínica, Universidade de Sao Paulo, Av. Prof. Lineu Prestes 748, Sao Paulo, SP 05508-900, Brazil
FEATURES
source
Location/Qualifiers
1..10275
/organism="Xanthomonas campestris pv. campestris str. ATCC 33913"
/strain="ATCC 33913"
/db_xref="ATCC:33913"
/db_xref="taxon:190485"
/note="pathovar: campestris"
470..559
/gene="XCC0657"
470..559
/gene="XCC0657"
CDS

/note="putative; ORF located using Glimmer/Genemark"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="AAM39973.1"
 /db_xref="GI:21111663"
 /translation="MRLSAPRAGTYIVKLDGASFDGVSLARQ"
 1092. .8147
 /gene="xada"
 /note="XCC0658"
 1092. .8147
 /gene="xada"
 /note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer/Genemark"
 /codon_start=1
 /transl_table=11
 /product="outer membrane protein"
 /protein_id="AAM39974.1"
 /db_xref="GI:21111664"
 /translation="MNRIRKYWNKSLGWAVASELASGDSPGSVASALIDRRQGLS
 LAAALALAGSAGIAIPLISAQSVSEVGRGASAPAHATAIGASNASATGAVATGAD
 SSASGVNSSAIGROTNAGENAVAIIGNSFVROAGENGVALGANAGVTGANSVALGAG
 SRTHEDDVSVSGNGRGGPATRRITNVSAVNANDAVNAOLOAVSEVAEDTATPEK
 AQPDDDSIGAYADGAGAVAGDAANAFSGSALAVSGANALADNATAMGANALAVGN
 SAAFGYNAQANGPGSVAVGVADEDGNPLITNGVPTDGTAGTAVGTVAGVASA
 DGFASSTYGVGAYAGAOSSAFGVSNATGDTYSTAVGTQTSAGTSSSTAVGVPDLIP
 GLGFEVQTOASGEAASALGAGIASGTYTTAVGTLSEAGTEATVAVGFAYAPGEGAT
 AVGESWASGELSTALGIYSTARGANSVALGANSVATRADTVSVGAAGERQITSVAA
 GTEGTDVNLNQLTAVSDVASATARSFVATGDGAIAEGDVAAAGSNAAFNDYSTA
 LGSSSVASAOGATAVSGSANGATTDNATAVGFNSTVAVENTTALGSSASGDSSTAVG
 GATRATASGATALGESIANGDSTALGVSVAFGCTVAVGASVAFGTDSSAARGN
 AAAGTASTAIGANSNAFGERTVALGGASNAGDESIALGVSSLASALGTTAVGSNAN
 ASIANATAVGNSAGDYPVALGSDSNASGYFTAVGCTSIANGRATAIAGESIGN
 GAASTALGFASVAMGEGGTAIGETSVAYGDDSTALGANAYAADTGSIAVGTIANAYCP
 RAISLGGOSNAGDESIALGWEAOEGDQIALGAGSQODAYSTAIGVATASGASAT
 AVGNNSRAVDGYATALGSDSNASGNESTTVGASVASGKGTATIGAESVARMDRTAI
 GTESVADGGDSTALGANARASYSVALGANANSNYVALGTYAVATIGSATSGC
 QSTAPGNESVALGWQSNASGTRSVSLGSGAYTPADDGYALGAGSIADRNTVSVSGV
 SERQITNVAAGTEGTDVAVNDOLNAVAGTAETRLVAGTGGDTAFAGGEDATAGSN
 ATADADYSSAFGASSQATATIGAVAIIGSASATAYANASGNAAGSAGSVSTGASQ
 ASGDYVALGGESEASGAQSTAVGAAGAAGSGDAFAGALSVAEGESTALGFASAT
 GESATAVGAESVADGTSAAFGFAGAEATSNYSTALGYSSTAGFNSTALGNFSTAGS
 SSYAVGGDATASGAYSTIAGASVASGYNSVAVGALLGLPTASGDEFTALGGAAW
 APLNSTALGNFAESTGESVALGDSVADRFVAVSVSGSNERQITNVAAGTOGTDVA
 VNLDQLTAVETAQGTSKYFRKSGSDSDAGATIEGDNALAAEGANASSDNSTAVGA
 GAOVAENATAVGMDALASGICAAALGNNAOLGENSNAVGNALASDIGATANGAGA
 QAISTYATALGSEAVASDNDATAVGFPSAASNVGSAARGYSESSGRLLSALGYAVA
 SSDYSTAVGAASIASGASAVAVGEFSEATGDESVAAGSGTFEEFIIPARASGTGAARF
 GGAMATADYTTAIGWNSYADGVNASALGQSAALADNALAIGNSRADAGASVAVGD
 ASATGINSTGVGRQVNVIGENAVSVGYNSFVRESAVNGVALGANAGATGADSVALGSG
 SRTYEANTVSVSGNGRGGPATRRIVNVSDGEVATDAVNGQDLADVQTTSGWQ
 TTGEGVATATGDRATAGAGATASGARSVAVAAGSTASATGASAMGVDSASGVNSA
 MGRQTNISGENGVALGYNSFVROAGENGVALGANAGASGDSVALGSGSRTYEANVS
 VSGSNGRGGPATRRIVNGACAIADASTDAINGGQFPQSIQNTASFLGGAAIGAQGV
 FVAPTYVIOGASYNNGAALTALDSKVTELDARGGTATATARTVSLRTAAVPAAVAGS
 ATVAASDVTDDSSAVQGTPTAAVGTATPATSTAVGNAVANHITGTAIGGSAYAHG
 PNDTAIGSNARVNADGSTAVAGANTQIAAVATNAVAMGEGAOVSAASGTAIQOGARSA
 OGAVALGQGSVADRANTVSVSGVGERQVANVAAGTRATDAVNKGQLDSGVAANSTYT
 DSRYSAMADSFETYQGDIEDRLRQRNRRLDROGAMGSAMLMNSASVAGIASQNRIGAG
 VGFQNGESALSVGYQRAISPRATVTIGGALSGDSSSIGVAGFGW"
 8306. .10186
 /gene="XCC0659"
 8306. .10186
 /gene="XCC0659"
 /note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer/Genemark"
 /codon_start=1
 /transl_table=11
 /product="protease"
 /protein_id="AAM39975.1"
 /db_xref="GI:21111665"
 /translation="MIDKNFRNLPLTGAILMALGASGTLVAAPALQVKEPTQADPAA
 SDFSSRLIVRYKDGTAASDRSSKLGVVQSAVGRAAPVTGARAGATAKATYLRKLGIT

BASE COUNT	1792 a	3332 c	3541 g	1610 t	
ORIGIN					
Alignment Scores:					
Pred. No.:					
Score:					
Percent Similarity:					
Best Local Similarity:					
Query Match:					
DB:					
	1		Gaps:	25	
US-09-825-414-7 (1-486) x AE012164 (1-10275)					
QY 5	ArgArgValGInGlnProProValThraAlaThrAspSerPheArgThr-AlaSerAsp-- 23				
Db 5870	CAGCGAATCCAGCGCGCCGCTGTCTGGCACATGGGCTACGGCGGAGTGGCGTCCAGCGA 5929				
QY 24	-----AlaSerLeuAlaSerSer----- 30				
Db 5930	CTACAGCACCGCAGTGGGTGCCCGCTGCTGGCCTCCGGCGCCAGCGCGGTGGCGGTGGC 5989				
QY 31	----ValArgSerValSerSerAspGlnGlnArgGluLeuAsn----- 43				
Db 5990	TGAATTACGGAAGCCACCGCGGATGAAGCGTGGCGGTGGCGGAGCACAGTCTTCGCG 6049				
QY 44	-----Al 44				
Db 6050	TTTCATCCCGGACGTCGTCGGGACCGGTCAGCGGCTTCGCTGGCGGTGCATGGGC 6109				
QY 44	AlleAlaAspTyrLeuThrAspHisValPheAlaAlaHisLysLeuProProAlaAspSe 64				
Db 6110	CACGGCCGATTAACACACCGCGGATCGGCTGAATTCTATGCCGACGGCGGTCAACGCCAG 6169				
QY 64	rAlaAspGlyGlnAlaAlaValAspVal-----HisAsnAl 76				
Db 6170	TGCATTGGCGCAGAGCGCGCCGATTCGCCGATTAACGCGCTGGCGCATCGCGGCAACAG 6229				
QY 76	aGlnIleThrAla-LeuIleGluThrArgAlaSer-----ArgLeuHisPheGluGly 94				
Db 6230	CCGCGCCGATGGCATTTGTCACAGCGTGGTCCGGCTGGATGCGTCCGCCACCGGCATCA 6289				
QY 94	LutThrProAlaThrIleAlaAspThrPheAlaLysAlaGluLysLeuAspArgLeuAlaT 114				
Db 6290	CAGCACCGCGCTGGCGCCCGCAGGTCAACGTAATCGCGAAACCGCGGCTCTCGGTGGCTA 6349				
QY 114	hrThrThrSerGlyAlaLeuArgAlaThrProPheAlaMetAlaSerLeuLeuGlnTyrM 134				
Db 6350	CAACTCGTTCTGCGCG--AGAGCGCGGTCAACGGTGTGCGTGGCTGCCAATGCCGG 6406				
QY 134	eGlnProAlaIleAsnLysGlyAspTrpLeuProAlaPro-----LeuLysProLeuT 152				
Db 6407	TGCCACCGCGCGCATTCGCTGGCATTCGGCTCCGGCTCGCGCATTAACGAAGCCAAC-A 6465				
QY 152	hrProLeuIleSerGlyAlaLeuSerGlyAlaMetAspGlnValGlyThrLysMetMet 172				
Db 6466	CGGTGTGCGTC--GGTAGCGGCAATGGCGCGCGCGCGGCGGCAACGCGTCCGATCGTCA 6522				
QY 172	spArgAlaThrGlyAspLeuHisTyrLeuSerAlaSerProAspArgLeuHisAspAlaM 192				
Db 6523	ATGTACGCGACGCGGCAAGTCGCCACCGATCGGTCACACAGGCGCACGCTG--GATGCAT 6579				
QY 192	eAlaAlaSerValLysArgHisSer-----ProSerLeuAla 205				
Db 192	eAlaAlaSerValLysArgHisSer-----ProSerLeuAla 205				

Db 6580 TGGCGCGCATGTGCAGACACCACCGCGCATGCTGCAGACACCACCGCGGCGCGGTGCCCC 6639

QY 205 rgGlValLeuAspThrGlyValAlaValGlnThrTyrSerAlaArgAsnAlaValArgT 225

Db 6640 GTGCACGCGCGCATCGCGCCACCAGCGCGCGGTGCCAGCCGCGCGGTGCACGCA 6699

QY 225 hrValLeuAlaProAlaLeuAlaSerArgProAlaValGlnGlyAlaValAspLeuGly 245

Db 6700 GCGTG---GCGGTAGCCGCGACGAGCAGCAGTCCCTCGGCCACCAGCGCGCATGGGTG 6756

QY 245 aL---SerMetAlaGlyGlyLeuAlaAlaAsnAla-----G 256

Db 6757 TGGACAGTAGCGCCACCGCGCGGTGACACAGCAGCGCGATGGGGCGCAGACCACTCCATTG 6816

QY 256 lypheGlyAsnArgLeuLeuSerValGlnSerArgAspHisGlnArgGly---GlyAlaL 275

Db 6817 GCGAGACGCGCGTGGCGGTGGGTACAACTCTTTGTGCGCCAGAGCGCGCGCCAACGCCG 6876

QY 275 euValLeuGlyLeuLysAspLysGluProLysAlaGlnLeuSerGluGluAsnAspTrpL 295

Db 6877 TCGCCCTCGGTGCCAATGCGGGTGCACAGCGCGCGCACTCGGTGCGCGTTCGGGT 6936

QY 295 euGluAlaTyrLysAla-----lLysSerAlaSerTyrSerGlyAlaAlaLeuAsnA 313

Db 6937 CGCGCACCTATGAAGCCAACGTGTGTGCGTGGGCGCAGCGCAACGCGCGGTGCCCCG 6996

QY 313 lAgLysArgMet-----AlaGlyLeuProLeuAspMetAlaThrAspAlam 329

Db 6997 CGACGCGTCGCATGCTCAATGTGGGTGCCCGCGCATTGCCGATGCCAGCAGCATGCGA 7056

QY 329 etGlyAla-----ValArgSerLeuValSerAlaSerSerLeuThrGlnAsnGlyL 346

Db 7057 TCAACGGCGGCCCACTTCTTCAGTGCCTCAGCAATACCGCCAGCTTCTCGCGGTGTG 7116

QY 346 euAlaLeuAla-----GlyGlyPheAlaGlyValGlyLysLeuGlnGluMetAlaThrL 364

Db 7117 CGGCGATCGGCGCGCAAGGTGTGTTCGTGGCGCGACCTACGTCATCCAGGTGCAGCT 7176

QY 364 ysAsnIleThrAspProAlaThrLysAlaAlaValSerGlnLeuThrAsnLeu----- 381

Db 7177 ACAACACAGTGGTCCCGCATTTGACCGCGTGGACAGCAAGTCAACCGAGCTGATGCAC 7236

QY 382 --AlaGlySerAlaAlaValPheAlaGlyTrpThr-----ThrAlaAlaLeuT 397

Db 7237 GTGGCGGACACCGCAGCGCTACCGTGCACGCAAGGTTTCGCTGCGCAGTGCAGCGGT- 7295

QY 397 hrThrAspProAlaValLysLys-----AlaGluSerPheIleGlnAspThrV 413

Db 7296 -----CCGCGGTGGCTGCATCCGCGACGGTCCGAGCAGCGATGTCTACTGACAGCA 7347

QY 413 alLysSerThrAlaSerSerThrThrGlyTyrValAlaAspGlnThrValLysLeuAlaL 433

Db 7348 GCGCAAGCGTGCAGGGCACGCCGACCGGTGGGTGTTGGC--ACGGCCACGCGCGCGG 7404

QY 433 ysThrValLysAspMetGlyGlyGluAlaIleThrHis-ThrGlyAlaSerLeuArgAsn 452

Db 7405 CGACCTCGACGCGCGTGGCAACGCCGTGTGGCCAACACATCACCAGCGCGCATTG 7464

QY 453 ThrValAsnAsnLeuArgGlnArgProAla-----ArgGluAlaAspIleGluGluGly 470

Db 7465 GTGGCAGCGCGTATGCGCATGGCCCCAACAGACACCGCGATTTGGCAGCAATGCACGCGTCA 7524

QY 471 GlyThrAlaAlaSerProSerGluIlePro-----PheArgPromet 484

Db 7525 ATGCCGATGCGACGACCGCGGTGGGTGCCAACAGCAGATTTGCTGCGGTGGCGACCAATG 7584

QY 485 Arg 485

Db 7585 CGG 7587

RESULT 9

AE000234

LOCUS AE000234 13431 bp DNA linear BCT 01-DEC-2000

DEFINITION Escherichia coli K12 MG1655 section 124 of 400 of the complete genome.

ACCESSION AE000234 U00096

VERSION AE000234.1 GI:1787633

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

1 (bases 1 to 13431)

Blattner,F.R., Plunkett,G. III, Bloch,C.A., Perna,N.T., Burland,V., Riley,M., Collado-Vides,J., Glasner,J.D., Rode,C.K., Mayhew,G.F., Gregor,J., Davis,N.W., Kirkpatrick,H.A., Goeden,M.A., Rose,D.J., Mau,B. and Shao,Y.

The complete genome sequence of Escherichia coli K-12

Science 277 (5331), 1453-1474 (1997)

97426617

9278503

2 (bases 1 to 13431)

Blattner,F.R.

Direct Submission

Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.

Email: ecollegenetics.wisc.edu Phone: 608-262-2534 Fax: 608-263-7459

3 (bases 1 to 13431)

Blattner,F.R.

Direct Submission

Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.

Email: ecollegenetics.wisc.edu Phone: 608-262-2534 Fax: 608-263-7459

4 (bases 1 to 13431)

Plunkett,G. III.

Direct Submission

Submitted (13-OCT-1998) Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA

This sequence was determined by the E. coli Genome Project at the University of Wisconsin-Madison (Frederick R. Blattner, director). Supported by NIH grants HG00301 and HG01428 (from the Human Genome Project and NCHGR). The entire sequence was independently determined from E. coli K12 strain MG1655. Predicted open reading frames were determined using Genemark software, kindly supplied by Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA, 30332 [e-mail: mark@amber.gatech.edu]. Open reading frames that have been correlated with genetic loci are being annotated with CG Site Nos., unique ID nos. for the genes in the E. coli Genetic Stock Center (CGSC) database at Yale University, kindly supplied by Mary Berlyn. A public version of the database is accessible (<http://cgsc.biology.yale.edu>). Annotation of the genome is an ongoing task whose goal is to make the genome sequence more useful by correlating it with other data. Comments to the authors are appreciated. Updated information will be available at the E. coli Genome Project's World Wide Web site (<http://www.genetics.wisc.edu>). *** The E. coli K12 sequence and its annotations are periodically updated; this is version M54. No sequence changes. Annotation updates: updated gene identifications and products; all new functional assignments courtesy of Monica Riley; added promoters, protein binding sites, and repeated sequences described in reference 1. The unique numeric identifiers beginning with a lowercase 'b' assigned to each gene (protein- or RNA-encoding) are now designated as gene synonyms instead of labels. This should allow them to be searched for in Entrez as gene names.

FEATURES

source

1..13431

location/Qualifiers

1..13431

organism="Escherichia coli K12"

strain="K12"

sub_strain="MG1655"

db_xref="taxon:83333"

<1..1145

note="ISS"

repeat_region

gene

complement(97..1077)

/ gene="trs5_5"
/ note="b1370"
complement(97.1077)
/ gene="trs5_5"
/ function="IS, phage, Tn; Transposon-related functions"
/ note="f326; 97 pct identical to Y152_ECOLI SW: P03837 but has 12 fewer N-terminal amino acids"
/ codon_start=1
/ transl_table=11
/ product="IS transposase"
/ protein_id="AAC74452.1"
/ db_xref="GI:1787634"
/ translation="MSHQLTFADSEFSTKRQRKEIFLSRMEQILPWONMTAVIEPFPKAGNGRRPYPLETMLRIHCQMHWYNLSDGAMEDALYEIASMRLEFARLSLBSALPDRTTIMNFRHLLEQHQIARQLFKETINRWLAEGVMTQGLVDATIEAPSSTKNEQQRDPEMHQTKGNQWHEFGMAKHIGVDKSGLTHSLVTTTAANEHDLNQLGNLKHGEQFVSADAGYOGAPQREELAEVDVWLLAERPGKVKTLKONPRKNKTAINIEYMKASIRARVEHPFRIRKQFGFVKARYKGLKNDQLAMLFTLANLFRVDQMIRQWERSQ"
709.736
/ note="factor Sigma70; predicted +1 start at 1426416"
729.758
/ note="factor Sigma70; predicted +1 start at 1426438"
874.1335
/ gene="b1371"
874.1335
/ gene="b1371"
/ function="orf; Unknown"
/ note="o153; This 153 aa ORF is 57 pct identical (0 gaps) to 66 residues of an approx. 208 aa protein VLOW_LAMB SW: P03701"
/ codon_start=1
/ transl_table=11
/ product="orf, hypothetical protein"
/ protein_id="AAC74453.1"
/ db_xref="GI:1787635"
/ translation="MLHAVNTQHGLOIRIGPSAIARLCIKRLDSDGHILPWQNLHAGEENLESGLTALSAEFTIGEGELMAHDVPLGAPDEYDDLISGCSHLPGKTHDVLGSDGRHSNTSLAWGAGVQFNPTESVAIDIAIEYEGSGDWRFTDGFIVGVGYKF"
complement(1230.1256)
/ note="factor Sigma70; predicted +1 start at 1426896"
1304.1330
/ gene="b1371"
/ note="factor Sigma70; predicted +1 start at 1427010"
complement(1346.1363)
/ note="central position to predicted promoter: -115"
/ bound_moiety="phob predicted site"
1394.4762
/ gene="b1372"
1394.4762
/ gene="b1372"
/ function="putative membrane; Not classified"
/ note="o1122; This 1122 aa ORF is 76 pct identical (1 gaps) to 381 residues of an approx. 408 aa protein Y401_LAMB SW: P03764"
/ codon_start=1
/ transl_table=11
/ product="putative membrane protein"
/ protein_id="AAC74454.1"
/ db_xref="GI:1787636"
/ translation="MNMAVKISGLKDGTKGPVONCTIQLAKRNSTTVVNTLASENPDEAGRYSMDEVYGGQSVILLVEGFPSPSHAGTIVYEDSQPTLNDFLGAMTEDDARPEALRRFELMVEEVARNASAVIAONTAAAKKSADASTSAREATHAADAADSARAASTSAGOASSAQSASSAGTASTKATEASKSAAAEESSKAATSGAAKTSETNASSLIQSAATSASTATTKASEAATSARDAASKEAKKSETNASSASSASSASATTAAGNSAKAAKTSETNARSETAAGOSAAAGSKTAAASSASASTSAGOSASATTAAGSAESASAASASTATTKAGEATEQASAAARSASAKTSETNAKASETSAESKSTAASATRAETAARSSASASKDEATROASAKSSATTSATKATEAAGSATYAAOSKSTAESATRAETAARAEADIASVALEDASTRKGIIVQLSSATNSTSETLAATPKAVKSAYDNAEKRLQKDQNGADIPDKGCFLLNNINAVSKTDFADKRGMRVVRVNAIPAGATSGKYPVVVMRSAGSVSELASRVITTTATRTAGDPMNCEFGNGFVMPGGWTDGRRAYGMFQYONNERAITHSIMSNKDDDLRSVFYVDCGAAPFEVAFIEDGLSISAPGADLVVNDTTYKGEATNPATECIAADVILDEKSGRGFYESHSLIVNDNLSCCKLFAFDEIYARGGNQIRMIGGAYGALWRNDGAKTYLLLTNOGDVYGGWNTLRFPAIDNATGELVIGTKLSASLNGNALTATKLOTPRRVS

/ gene="ynac"
/ note="b1373"
4762.5337
/ gene="ynac"
/ function="orf; Unknown"
/ note="o191; This 191 aa ORF is 94 pct identical (0 gaps) to 191 residues of an approx. 200 aa protein Y194_LAMB SW: P03740"
/ codon_start=1
/ transl_table=11
/ product="orf, hypothetical protein"
/ protein_id="AAC74455.1"
/ db_xref="GI:1787637"
/ translation="MAFRMSEQPRTIKYINLAGTNEFIGEADAIYPHTGLPANSTDIAPDIPAGFVAVFNSDESSWHLVEDHGRKTVYDVASGNALFISELGPLPENVTLWSPEGEFQKWNGTAWVKDTEAEKLFRIREAETKNNLMQVASEHIAIPLQDAADLEIATEEEISLEAWKXYRVLLNRVDSTQAODIEWPALP"
complement(5435.6025)
/ gene="b1374"
complement(5435.6025)
/ gene="b1374"
/ function="IS, phage, Tn; Not classified"
/ note="f196; This 196 aa ORF is 76 pct identical (0 gaps) to 193 residues of an approx. 200 aa protein TNP0_ECOLI SW: P05823"
/ codon_start=1
/ transl_table=11
/ product="putative transposon resolvase"
/ protein_id="AAC74456.1"
/ db_xref="GI:1787638"
/ translation="MSRIFAYCRISILDQTTENQREIESAGFKIKPOQIIIEHISGSAATSERPGFNRLIARLKCQDQIVTKDLRGCNAMDIRKTEVQLTEGIRVHCLALGIDLTSPTKMMQVYISAVAEFERDILLERTHSGIVRARGAGKRFGRPVNLNEQKQVFERIKSGVSISAIAREFKTSROTILRAKAKIQTPI"
complement(6033.6061)
/ note="factor Sigma70; predicted +1 start at 1431699"
complement(6342.6608)
/ gene="ynae"
/ note="b1375"
complement(6342.6608)
/ gene="ynae"
/ function="orf; Unknown"
/ note="f88; This 88 aa ORF is 32 pct identical (4 gaps) to 71 residues of an approx. 1112 aa protein SVY_YEAST SW: P07806"
/ codon_start=1

Alignment Scores:
Pred. No.:
Score:
Percent Similarity:
Best Local Similarity:
Query Match:
DB:
US-09-825-414-7 (1-486) x AE000234 (1-13431)
Oy 19 ArgThrAlaSerAspAlaSerLeuAlaSerSerValArgSerValSerSerAspGln 38
111 111111 ::::::::::: ::::11111
Db 1742 CGTAACGCCGTCGCCGTGCACAGACACGCCGCGGAGAGAAAGTCAGCCAGTGAAGCC 1801
Oy 39 GlnArgGluIleAsnAlaIleAlaAspTyrIleuThrAspHisValPheAlaAlaHisLys 58
111 :::: 111 11111
Db 1802 AGCACATCAGCCCGTGAGCGCGGCAACCCATGCGGCTGAT-----GCTGCGGACTCA 1852
Oy 59 LeuProProAlaAspSerAlaAspGlyGlnAlaAlaValAlaAspValHisAsnAlaGlnIle 78

Db 1853 GCACGGCAGCCAGCAGCTCAGCCGGACAGCGCGCTCGTCGGCTCAGTCAGCGTCTTCC 1912
QY 79 ThrAlaLeuILeGIuThrArgAlaSerArgLeuHisPheGIuGIuThrProAlaThr 98
Db 1913 AGCGCA-----GGACGGCATCAACAAGGCCACT 1942
QY 99 IleAlaasp---ThrpheAlaLysAlaGIuLysLeuAspArgLeuAlaThrThrThrSer 117
Db 1943 GAAGCATCAAAAAGTGTGCTGCCCTGCAGAGTCTCAAAAAGCGCGCGCCACAGTGC 2002
QY 118 GlyAlaLeuArgAlaThrProPheAlaMetAlaSerLeuLeuGIuTyrMetGIuProAla 137
Db 2003 GGTGGCGCAAAACGTCAAGAAAGAATGCTTCAGCGTCACTACAATCAGCAGCCACATCT 2062
QY 138 IleasnLysGIuAspTyrPleuProAlaProLeuLysProLeuThrProLeuIleSerGIy 157
Db 2063 GCATCCACCGCG-----ACCACGAAGGCATCAGAA 2092
QY 158 AlaLeuSerGIyAlaMetAspGIuValGIyThrLysMetMetAspArgAlaThrGIyAsp 177
Db 2093 GCTGGACCTCGGCGCCGGATGGCGCGCTCAAAAAGAGCGGCAAAATCATCA--GAA 2149
QY 178 LeuHisTyrLeuSerAlaSerProAspArgLeuHisAspAlaMetAlaAala-----Ser 195
Db 2150 ACGAAGCATCATCAAGCGCCAGTAGTCAGCTTCTCGGCAACGGCGGAGAAATTC 2209
QY 196 ValLys-----ArgHisSerProSerLeuAlaArgGIu 206
Db 2210 GCGAAGCGGCAAAAACGTCCGAGACGAACGCCAGGTCTTCTGAACGGCAGCGGAGCAG 2269
QY 207 ValLeuAspThrGIyValAlaValGIuThr-----TyrSerAlaArgAsnAlaValArg 224
Db 2270 AGCGCTCGGCTCGGCGAGGCTCAAAAACAGCGGCTGCTGCTGCTGCAGTGCAGCTCA 2329
QY 225 ThrValLeuAlaProAlaLeuAlaSerArgProAlaValGIuGIyAlaValAspLeuGIy 244
Db 2330 ACAAGTCCGGGCGAGCGCTCAGCCAGTCCACCGCGCGGAAATCGGCAGAAAGCGCC 2389
QY 245 ValSerMetAlaGIyGIyLeuAlaAlaAsnAlaGIyPheGIyAsnArgLeuLeuSerVal 264
Db 2390 GCATCGTCTGCTTCAACAGCCACAACGAAGGCTGGCGAAGCCACTGAACAGGCCAGCGCA 2449
QY 265 GluSerArgAspHisGluArgGIyGIyAlaLeuValLeuGIyLeuLysAspLysGIuPro 284
Db 2450 GCAGCGAGGTCT-----GCTTCCGCGACGGAAGACATCCGAACG 2488
QY 285 LysAlaGIuLeuSerGIuGIuAsnAspTyrPleuGIuAlaTyrLysAlaIleLysSerAla 304
Db 2489 AACGCGAAAGCGTCGGAAACAAGC-----GCAGAATCTCAAAAACGGCTGCCGCATCG 2542
QY 305 SerTyrSerGIyAlaAlaLeuAsnAlaGIyLysArgMetAlaGIyLeuProLeuAspMet 324
Db 2543 TCAGCCAGTTTCGGCGGCTCATCGGCATCATCG-----GCGTCTGCTTCAAAAAGATGAG 2596
QY 325 AlaThrAspAlaMetGIyAlaValArgSerLeuValSerAlaSerSerLeuThrGluAsn 344
Db 2597 GCGACCAAGACAAGCGTCAGACGAAGAGC-----AGCGCACGACGCGCATCCACGAAG 2650
QY 345 GlyLeuAlaLeuAlaGIyGIyPheAlaGIyValGIyLysLeuGIuMetAlaThrLys 364
Db 2651 GCGACAGAGGCTGTGGCAGT----- 2671
QY 365 AsnIleThrAspProAlaThrLysAlaAlaValSerGIuLeuThrAsnLeuAlaGIySer 384
Db 2672 -----GCGACGGCGGACGCTCAGAGCAAAAGTACG-----GCGGAATTC 2710
QY 385 AlaAlaValPheAlaGIyTyrThrThrAlaAlaLeuThrThrAspProAlaValLysLys 404
Db 2711 GCGGCAACGGCGCGC-----GAGACAGCAGCTAAACGG 2743
QY 405 AlaGIuSerPhe-----IleGIuAsp-----ThrValLys----- 414

Db 2744 GCAGAGGATATTGCATCCGCCGTGGCGCTTGAGGATGCAAGTACGAGAAAAAGGGATA 2803
QY 415 -----SerThrAlaSerSerThrThrGIyTyrValAlaAspGIuThrValLysLeu 431
Db 2804 GTACAGCTCAGCAGTGCAGCCACACAGTACGTCTGAACGCTGGCGCAACGCCAAGGCA 2863
QY 432 AlaLysThrVal-----LysAspMetGIyGIuAla 442
Db 2864 GTAATAATCAGCCTATGACAATGCAGAAACGTCGCAAGAAAGACAGACGGCGGTGAT 2923
QY 443 IleThrHisThrGIyAlaSerLeuArgAsnThrValAsnAsnLeuArg----- 458
Db 2924 ATACCCGATAAGGATGCTTCCTG---ACAACATTAAACGGGTCAGTAAACAGACTTT 2980
QY 459 -----GluArgProAlaArgGIuAlaAspIleGIuGIuGIyGIyThrAlaAlaSerPro 476
Db 2981 GGTGATAAGCGGTGATAGCGTTATGTGCGGGTTAACGCTCTCGAGGTGCAACATCTGGA 3040
QY 477 SerGIuIleProPheArgProMetArgSer 486
Db 3041 AAATATTACCTGTGTGTATTATGCGTCT 3070
RESULT 10
AE007019 15424 bp DNA linear BCF 27-APR-2001
LOCUS Mycobacterium tuberculosis CDC1551, section 105 of 280 of the complete genome.
ACCESSION AE007019 AE000516
VERSION AE007019.1 GI:13881105
KEYWORDS
SOURCE Mycobacterium tuberculosis CDC1551.
ORGANISM Mycobacterium tuberculosis CDC1551.
REFERENCE 1 (bases 1 to 15424)
AUTHORS Fleischmann,R.D., Alland,D., Eisen,J.A., Carpenter,L., White,O., Peterson,J., DeBoy,R., Dodson,R., Gwinn,M., Haft,D., Hickey,E., Kolonay,J.F., Nelson,W.C., Umayam,L.A., Ermolaeva,M., Salzberg,S.L., Delcher,A., Utterback,T., Weidman,J., Khoult,H., Gill,J., Mikula,A. and Bishai,W.
TITLE Whole genome comparison of Mycobacterium tuberculosis clinical and laboratory strains
JOURNAL unpublished
REFERENCE 2 (bases 1 to 15424)
AUTHORS Fleischmann,R.D., Alland,D., Eisen,J.A., Carpenter,L., White,O., Peterson,J., DeBoy,R., Dodson,R., Gwinn,M., Haft,D., Hickey,E., Kolonay,J.F., Nelson,W.C., Umayam,L.A., Ermolaeva,M., Salzberg,S.L., Delcher,A., Utterback,T., Weidman,J., Khoult,H., Gill,J., Mikula,A. and Bishai,W.
TITLE Direct Submission
JOURNAL Submitted (25-APR-2001) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA
FEATURES
source location/Qualifiers
1..15424
/organism="Mycobacterium tuberculosis CDC1551"
/strain="CDC1551"
/db_xref="taxon:83331"
/note="clinical strain"
complement(171..581)
/gene="MT1491"
complement(171..581)
/gene="MT1491"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK45754.1"
/db_xref="GI:13881106"
/translacion="MTVMADRSGRPAVRRMKTLTQALNADKVEQVEDVLDGLGK TMAELNLSLSQLNSTVERLEDGLDLEGTLSLDLAKRLIYVEPVEAIVRIDYIV SLGETVMSPLSTEHAVRGVLDRLRNRTVHEPTN"
complement(598..1341)
gene

CDS /gene="MT1492" complement(598. .1341) /gene="MT1492" /note="similar to GB:U00013 PID:466869; identified by sequence similarity; putative" /codon_start=1 /transl_table=11 /product="glucosamine-6-phosphate isomerase" /protein_id="AAK45755.1" /db_xref="GI:13881107" /translation="MSSSIEIFPDSILVAAAGKRLVGAIGAAVAARGQALIVLTGGG NGIALRLYLSAQAOIEMWSKVLFWGDERYVPEDDERNLKQARRALLNHVDIPSNQY HPMASDDEFGDLDAALAYEQVLASAAPGDPAPNFDVHLGEGEGHINSLFPHS PAVLESTRMVAVDSPKPPRRITLILPAIQRSREWLLVSGPGKADAVAAAIIGAD PVSVPAGVVGQNTLMLLDRDAAAKLPS" complement(1338. .2249) /gene="MT1493" complement(1338. .2249) /note="similar to GP:5459405; identified by sequence similarity; putative" /codon_start=1 /transl_table=11 /product="oxpocycle protein OpcA" /protein_id="AAK45756.1" /db_xref="GI:13881108" /translation="MIVDLPDTTAVNKKLDELREKIGAVAMGRVLTIIAPDSEAM LEESIEANDASHEPSRIIVTMRGDYADRLDAQLRVGADAGGEFVLRISGPL AGHADSIVIPLLPDIPVAVAMPDIAPAPQDALGKLAIRITDATNAIDPLSAIKS RLAGYAGDTDLAWSRITYRATLTSAYDQPPHEPIESALVSGLKEPALDVLAWLA SRIEGPVRAVVELKVELVRNSETIVLSRPQEGITATLTRIGSKPALVPLARRVTGEC LAEDLRLLDPDEIYCALEGIKKVQYR" complement(2302. .3846) /gene="MT1494" complement(2302. .3846) /note="similar to GP:5459404; identified by sequence similarity; putative" /codon_start=1 /transl_table=11 /product="glucose-6-phosphate 1-dehydrogenase" /protein_id="AAK45757.1" /db_xref="GI:13881109" /translation="MKPAHAAASWRNPLRDKRKLPRIAGPCGVIFGVTDLARKK VMPAVYDIANRGLLPFTSLVGFAARDWSTQDFGQVVYNAVQEHCRTPFRQNMRLA EGFREVPGTFEDDDAFAQLAETLEKDAERGTGNGHAFYLAIPKSEFPVYCEQLHKS G LARPQGRWSRVVIEKPEFGHDIASARELNKAVNAVPEEAVFRIDHYLGKETVONILA LRFANQLEDPIWNAHYVDHVQITMAEDIGLGRAGYYDGIIGARVYIQNHLMQLALT AMEEPVSFHPALQAEKIKVLSATRLAEPILDQTTSRGOYAAWGQGEKVVGLDEEGF AEDSTTEFFAAILTEVDTRWAGVPFYLTRGKRLGRVTEIALVFRRAPHLPFDATMT DELGTNNAVIRVOPDEGVTLRFGSKVPGTAMEVRDVNMFSYGSAFEDSPEAYERLI LDVLLGEPSLFPVNAEVELAMEILDPALHWAHAGTPDAYEAGTWGPRESSLEMLRRTG REWRRP" complement(3843. .4964) /gene="MT1495" complement(3843. .4964) /gene="MT1495" /note="similar to GP:2398718; identified by sequence similarity; putative" /codon_start=1 /transl_table=11 /product="transaldolase" /protein_id="AAK45758.1" /db_xref="GI:13881110" /translation="MTAQNPNLAAISAAGVSVWLDDLSDRLRSGNLQELIDTKSVVG VTTNPSTQKALSEGHTYDAQIAELAAAGADVATIRVTPTDVRSACDVLVPQWEDS DGVDRVSTIEVDPRLAHETEKTIQQAIELMKIVDRPNLFIKIPATKAGLPASIVLAE GISVNVTLIFSVOYRREVMDAYLTGMEKARQAGHSLSKIHSVASFFVSRVDTEIDKRL DRIGSRQALELRGQAGVANARLAAYAREVEEDSDRYRSLKVDAARVQRPLMASTGVK NPDYSDTLVYTELVAPHTVNTIMEKTIIDAVADHGVIOGDVTVTASDAQAVFDQLGAI GIDLTFVFAVLEEGVRKFEASWNEILQETRAHLDTAAQ" complement(4981. .7158) /gene="MT1496" complement(4981. .7158)

CDS complement(4981. .7158) /gene="MT1496" /note="similar to GB:U00013 SP:P46708 PID:466894 PID:2398717; identified by sequence similarity; putative" /codon_start=1 /transl_table=11 /product="transketolase" /protein_id="AAK45759.1" /db_xref="GI:13881111" /translation="MRHCKRRDSRRRVRTDHLADSRTFEVYTLLEISALTRPRHPDM TEIDSAVDTIRVLADAVQKYNCGHPGTAMSLAPLATYLFQRTMRHDPSTDHWTGRLD REVLSAGHSSLTYIQLYLGFGLELSDIESLRTWGSKPPGHPEFRHTPGVEITTGPL GQGLASAVGMAMASRYERGLFDPDAEPGAPFDHYIYVLAASDGDIEEGVTSASSLAA VQQLGNLIVFYDRNQISTIEDNTALCEDTARYARYGMVQVEVEGENVGIIEATA NAOAYTDRPSFIALRTVIGYPAPNMDTGKAHGAALGDDEVAAVKKIVGFDPKTFQY REDVLTHTRGLVARGKAHERMOLEFDMARREPERKALLDRLLAOKLPDGMADLPY WEPGSKALATRAASGAVLSALGPRIPELWGSADLAGSNNTTIKGADSFGPSISTKE YTAHWYGRTLHFGVREHAMGAISGIVLHGPTRAYGTFIQFSDYMRPAVRLAALMDI DTIVWTHDSIGLEDGPTHQPIEHLUSALRAIPRLSVYRADANETAYAMRTILARRN GSGPYGLLTRQGVPLDGTDAEGVARGGYVLSADAGLQDGEEDVILIAATGSEVOILA VAAQTLIADNDILARVSMPCLEMEFEAQPYEYRDVILPPIYSAVAVEAGVAQCWHQL VGDTEIVSIEHYGESADHKTFRFXYGTAEVAAAAERALDN" complement(7521. .11747) /gene="MT1497.1" complement(7521. .11747) /gene="MT1497.1" /note="similar to GB:AL123456; identified by sequence similarity; putative" /codon_start=1 /transl_table=11 /product="pe_pgrs family protein" /protein_id="AAK45760.1" /db_xref="GI:13881112" /translation="MSLVIYADETVAAAAALDVARIGSSIGCANAAAGSTTSVLAAG ADEVSAIATLEGSAREYQAISTQVAAFHDRFAQTLTSAVGSYVSAETNAAPLATL EHNVLNALNAPTQALLGRPLIDGACAGAPGTQAGAGAGGILLWNGAGSGSAGPGQYGG AGGAGLEFGTGAGGAGGAGAGAGAGSGGWLNGVYAGAGQSLLGATGAGAGNA GLFVGGTGXPGPGPGVGGTGGAGLGGLTYGAGGHGAGGPGPIGGVGHGQYV GAAGLLGVGGHGGAGGHAEGVACAAGEDLSPHGTSGVGGADAGDGTGGRGWLIIAGA GGAGGAGVGTGGAGGAGFSRALIYAGDNGDGGNGMGAGGAGGAPGAGGLISL GGQAGGAGGTGGAGGVGGDGGAGGPGNQAFNAGAGGAGHGDDPGAGGAGGTGAGS TIGAHPSAGASPTSGGAGAGNGAHFSSSGKAGAGGAGGAGGLVNGGAGGAGNG APGAPPSGDDPNNGGGGAGGAGGKGGDGAQAAGDGGAGGAGGKGGNGNGATGATGLN GIGAGADGTDGKGKNGAGGGGAGGQGGKALAATHQDSMGAGGAGNGAGAGMGG DGGNGAKGTFDNGDGVGGNGGSGRGIGGAGGIGGAGSTAGADARGATPTSGNG GTGNGANATVAGGAGGASGKGNGLVNGAGGKGGDMAGVAGSSPTTAGESGTS GONGGAGGAGGAGRGDFGDDGTGGAGNGGADGAGGAGGAGANATTPGAKGDDG HGPGAGCGNGGGQGGPGGLAGNLFQONGIQGVGSSGKGGAGGLAGDGGNANGNPAF GDNGSGHCGNCGNPGAGGQGGSCGAGSTPGAKGAFPTPSGDDGDDGNGNSOYVG NGNGDGGNGNGSGSAGTGGNGRGGDGAFGMSANATNPGENPNPNNPGNGAGGAG GAGLNGNGGAGGNGGLGGFGGNGAAGANGVAVGAPGQPGGAGGAGGAGNGAGNG GQGVYSDGAGGAGGAGGDPAGDAGANGNGQAGAFAGGGGGRGGDGGNAGNAGAGG PGGTGSTAGKAGPAGSIXHDGNGGHGGAASGNGCPGHGNGGNGGTGANANG GIGGTGAGSTGAKGLGTNEGDGDDGGRGNGGRGNGOGLTGAGGNGGTGTPGN CGNGNGASGDLVTS PGDGGGGGRGDAGRGDAGLIGSSSPGPTPGDWGTGTTGTG GTGQGANGLTGGRGGTGNGGNGNTGTGGAGGTGTTGHNQSQPGMGNGGAGGFG GNGFAGVGGRGGMGSGGTGTTDAGPFGTGTGTTGGHGGQGGGGFSILLGLGLGG LGSPSITATGTAGGAGGGGFGGCGGEFV" complement(11938. .12096) /gene="MT1497.2" complement(11938. .12096) /gene="MT1497.2" /note="identified by glimmer2; putative" /codon_start=1 /transl_table=11 /product="hypothetical protein" /protein_id="AAK45761.1" /db_xref="GI:13881113" /translation="MVDKVTYAAEPQGRFTRLRHHIGTARQPEPRSGQFSHPDENRY LGLTISVTP" 12146. .13072 /gene="MT1498" 12146. .13072

AUTHORS Alba,H., Baba,T., Fujita,K., Hayashi,K., Honjo,A., Horiuchi,T., Ikemoto,K., Inada,T., Isono,K., Isono,S., Itoh,T., Kanai,K., Kasai,H., Kashimoto,K., Kim,S., Kimura,S., Kitagawa,M., Kitakawa,M., Makino,K., Masuda,S., Miki,T., Mizobuchi,K., Mori,H., Motomura,K., Nakamura,Y., Nashimoto,H., Nishio,Y., Oshima,T., Saito,N., Sampei,G., Seki,Y., Tagami,H., Takemoto,K., Wada,C., Yamamoto,Y. and Yano,M.

TITLE The systematic sequencing of the Escherichia coli genome in Japan

JOURNAL Unpublished

REFERENCE 3 . (bases 1 to 16902)

AUTHORS Mori,H.

TITLE Direct Submission

JOURNAL Submitted (14-DEC-1996) Hirotsada Mori, NARA Institute of Science and Technology, Res. & Edu. Center for Genetic Info.; 8916-5 Takayama, Ikoma, Nara 630-01, Japan

COMMENT (E-mail:hmori@gtc.aist-nara.ac.jp, Tel:81-7437-2-5660, Fax:81-7437-2-5669)

Collaboration Information:

Project:

The Japan E.coli genome DNA sequencing project

Group:

The Japan E.coli genome DNA sequencing group

Members: (1995.4 - 1996.3)

Alba,H., Baba,T., Fujita,K., Hayashi,K., Honjo,A., Horiuchi,T., Ikemoto,K., Inada,T., Isono,K., Isono,S., Itoh,T., Kanai,K., Kasai,H., Kashimoto,K., Kim,S., Kimura,S., Kitagawa,M., Kitakawa,M., Makino,K., Masuda,S., Miki,T., Mizobuchi,K., Mori,H., Motomura,K., Nakamura,Y., Nashimoto,H., Nishio,Y., Oshima,T., Saito,N., Sampei,G., Seki,Y., Tagami,H., Takemoto,K., Wada,C., Yamamoto,Y. and Yano,M.

Headed by:

Name: Takashi Horiuchi

Address: National Institute of Basic Biology, Okazaki, 444, Japan

E-mail: kishori@nibb.ac.jp

Information operator:

Name: Hirotsada Mori

Address: NARA Institute of Science and Technology, Ikoma, 630-01, Japan

E-mail: hmori@gtc.aist-nara.ac.jp

URL:

The Japan E. coli genome database

http:bsw3.aist-nara.ac.jp.

Location/Qualifiers

1. .16902

/organism="Escherichia coli"

/strain="K12"

/db_xref="taxon:562"

/map="30.7 min"

/clone="Kohara clone #264"

/clone_lib="Kohara lambda miniset library"

/note="Nucleotide position 1428466-1445367 from the initiation site of Thra (0 min.).~This clone is from kohara lambda miniset library"

813. .1841

/note="ORF_ID:o263#17

similar to [SwissProt Accession Number P03736]"

/codon_start=1

/transl_table=11

/product="Minor tail protein precursor H."

/protein_id="BAA14970.1"

/db_xref="GI:1742238"

/translation="MAEQTSRLAIIDSTGAKNADNLTSLVKMTQAGETAANSAGKVTKATEDEKNALAKLKAIDPVGAIDTVGRRYSELKFFDKGLIDKEEYEFVLRKLNETTEELSGVAQAQREAEKAGLIAAQEAQACAFQRMIDKITPLAALRNLEQOHELNAFASGKINGSQFENYSRKIQETRRELTEGAQAREAKAHDEQVALQRLTAOLDPVGTAFNRILVEQOKOLNEAKAKGMLSPMEYELSGLRAMRSELSEVLOSLSKTGMSAKQTAFAMRMLPAQMTDIVVGLSTGSPFVMVLMQGGHSAMQETRIASLPVTAIPAGMKIPCVTAGSA"

1817. .1972

/note="ORF_ID:o263#18

similar to [SwissProt Accession Number P03701]"

/codon_start=1

gene

/transl_table=11

/product="Outer membrane protein lom precursor (ORF 206A)."

/protein_id="BAA14971.1"

/db_xref="GI:1742239"

/translation="MNRNWFSSVMVGPVSVRVNWFSAAYAMAGMAYSRYSTFSGDYLRYTDNKRGE"

complement(2105. .3115)

/gene="IS5"

complement(2105. .>3115)

/gene="IS5"

/note="ORF_ID:o263#19

similar to [SwissProt Accession Number P03837]

start codon is not identified yet"

/codon_start=1

/transl_table=11

/protein_id="BAA14972.1"

/db_xref="GI:1742240"

/translation="FIRSAQORDIMSHQITPADSEFSTKRROTREKIFLSRMEQILPWQMTAVIEPFYPRKAGNGRRPYLETMLRIHCQHWYNLSDGAMEDALYEIASMRLEFARISLDSALPDRRTIMNFRHLLQHQLARQLFTINRWLAEGVAMTQGLVDATIIEARSSTKNKEQQRDPREMOTKKNQWGHGMAHIGVDAKSGLTHSLVTTAAENEDLNQGNLHGEEQFVSADAGYOGAPQREBELAEVDVDWLIAERPQVKYTLKQNPKNKTAINIEYMKASIRARVEHPFRILIKRQGFVKARYKGLKNDQNLAMFLTLANLFRVQDMIRQWER SQ"

2441. .2809

/note="ORF_ID:o263#20

similar to [SwissProt Accession Number P03838]"

/codon_start=1

/transl_table=11

/protein_id="BAA14973.1"

/db_xref="GI:1742241"

/translation="MLSMKQITQLIEVMLVGRGDQAVGQATLGIDTNVGLHAKVPLIAFLGLMHLRIALLFVLGRAGCLNDGIGHQSAIGHHDACFGQPAIDGLQLTGLMLLEQVAEIHHDGCAIROGAIQG"

2882. .3343

/note="ORF_ID:o263#21

similar to [SwissProt Accession Number P03701]"

/codon_start=1

/transl_table=11

/product="Outer membrane protein lom precursor (ORF 206A)."

/protein_id="BAA14974.1"

/db_xref="GI:1742242"

/translation="MDHAVNTOHGLQIRIGPSAIRLGIKRLDDSGHILPQONLLHAGEENLFSGLTALSAEFTIGEGELMAHDVPLGCAPDEYDDLISGTCSHLPKRTHDVLTGSDDGRHSNTSLAWGAVQFNPETSEVAIDIAEYEGSGDWRTDGFIVGVGYKF"

3408. .6770

/gene="37"

3408. .6770

/gene="37"

/note="ORF_ID:o263#22

similar to [SwissProt Accession Number P03744]"

/codon_start=1

/transl_table=11

/product="Tail fiber protein GP37."

/protein_id="BAA14975.1"

/db_xref="GI:1742243"

/translation="MAVKISGVLKDGTEGKPVQNCITQLKAKRNSTVVVNTLASENPD EAGRYSMDVEYGYQSVLLVEGFPPSHAGTITVEYEDSQPTLNDFLGAMTEDDARPEALRRFELMVEEVARNASAVAQNTAAAKKSSADASTSAREATHAADAADSARAASTSAGQAASSAQSSASSAGTASTKATEASKSAASAAEASSKSAATAGATSETNASSLSQSAATASATATTKASEAATTSARDAASKEAAKSSETNASSSASSAASATAGNSAKAAKTSETNARSSSETAAGQSSAAGSKTTAAASSASASTSAGQASASATAGKSAESAASSASTATTKAGEATEQASAAARSASAAKTSETNPAKASETSSESSKTTAAASSASASSASASKDEATROASAKSSATTAATKATEAAGSATAAQSKSTAESAATRAETAAKRAE DIASAVALEDAESTTKKGIQVLSATNSTSETLAATPKAVKSAYDNAEKRLQKDONGADIPDKGCFLLNNINAVSKTDFADKRGMRVVRVNPAGATSGKYYPVVVMRSAGSVSELASRVILTTATRTAGDPMNCCENFGFVMPGWTDRGRYAYGMFWOYQNNERRAIHISIMSNKGDDLRSVFYVDGAAPVFAFIEDGLSISAPGADLVVNDTTYKFGATNPATPCIAADVI LDFKSGRGFEYESHLLVNDNLSEKSLFATDEIYARGGNQIRMIGEGYCALMRNDGAKTYLLLTNOGDVYGGWNTLRPFAIDNATGELVIGTKUSASLNGNALTTATKLQTPRRVSGVEFDGSKDITLTAHVAFAFARRATDTYADADGCVWNAESGAYNVTNRSGDSYILVNFYT

D90774 LOCUS 18700 bp DNA linear BCT 29-MAY-1997
DEFINITION E.coli genomic DNA, Kohara clone #263(30.5-30.9 min.).
ACCESSION D90774 AB001340
VERSION D90774.1 GI:1742217
KEYWORDS Complete and shotgun sequencing; -37; IS5; racc; recE; recT; sieB; trkG; ydaE; ydaF.
SOURCE Escherichia coli (strain:K12) DNA, clone_lib:Kohara lambda miniset library clone:Kohara clone #263.
ORGANISM Escherichia coli
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.
REFERENCE 1 (sites)
Aiba,H., Baba,T., Fujita,K., Hayashi,K., Inada,T., Isono,K., Itoh,T., Kasai,H., Kashimoto,K., Kimura,S., Kitakawa,M., Kitagawa,M., Makino,K., Miki,T., Mizobuchi,K., Mori,H., Mori,T., Motomura,K., Nakade,S., Nakamura,Y., Nashimoto,H., Nishio,Y., Oshima,T., Saito,N., Sempel,G., Seki,Y., Sivasundaram,S., Tagami,H., Takeda,J., Takemoto,K., Takeuchi,Y., Wada,C., Yamamoto,Y. and Horiuchi,T.
A 570-kb DNA sequence of the Escherichia coli K-12 genome corresponding to the 28.0-40.1 min region on the linkage map
DNA Res. 3 (6), 363-377 (1996)
TITLE 2 (sites)
JOURNAL Alba,H., Baba,T., Fujita,K., Hayashi,K., Honjo,A., Horiuchi,T., Ikemoto,K., Inada,T., Isono,K., Isono,S., Itoh,T., Kanai,K., Kasai,H., Kashimoto,K., Kim,S., Kimura,S., Kitagawa,M., Kitakawa,M., Makino,K., Masuda,S., Miki,T., Mizobuchi,K., Mori,H., Motomura,K., Nakamura,Y., Nashimoto,H., Nishio,Y., Oshima,T., Saito,N., Sempel,G., Seki,Y., Tagami,H., Takemoto,K., Wada,C., Yamamoto,Y. and Yano,M.
MEDLINE The systematic sequencing of the Escherichia coli genome in Japan
REFERENCE 3 (bases 1 to 18700)
AUTHORS Mori,H.
TITLE Direct Submission
JOURNAL Submitted (14-DEC-1996) Hirotsada Mori, NARA Institute of Science and Technology, Res. & Edu. Center for Genetic Info.; 8916-5 Takayama, Ikoma, Nara 630-01, Japan
COMMENT (E-mail:hmori@tc.aist-nara.ac.jp, Tel:81-7437-2-5660, Fax:81-7437-2-5669)
Collaboration Information:
Project:
The Japan E.coli genome DNA sequencing project
Group:
The Japan E.coli genome DNA sequencing group
Members: (1995.4 - 1996.3)
Aiba,H., Baba,T., Fujita,K., Hayashi,K., Honjo,A., Horiuchi,T., Ikemoto,K., Inada,T., Isono,K., Isono,S., Itoh,T., Kanai,K., Kasai,H., Kashimoto,K., Kim,S., Kimura,S., Kitagawa,M., Kitakawa,M., Makino,K., Masuda,S., Miki,T., Mizobuchi,K., Mori,H., Motomura,K., Nakamura,Y., Nashimoto,H., Nishio,Y., Oshima,T., Saito,N., Sempel,G., Seki,Y., Tagami,H., Takemoto,K., Wada,C., Yamamoto,Y. and Yano,M.
Headed by:
Name: Takashi Horiuchi
Address: National Institute of Basic Biology, Okazaki, 444, Japan
E-mail: kishori@nibb.ac.jp
Information operator:
Name: Hirotsada Mori
Address: NARA Institute of Science and Technology, Ikoma, 630-01, Japan
E-mail: hmori@tc.aist-nara.ac.jp
URL:
The Japan E. coli genome database
http:bsw3.aist-nara.ac.jp.
location/Qualifiers
1. 18700
/organism="Escherichia coli"
/strain="K12"
/db_xref="taxon:562"

/map="30.5 min"
/clone="Kohara clone #263"
/clone_lib="Kohara lambda miniset library"
/note="Nucleotide position 1416913-1435612 from the initiation site of ThrA (0 min.).-This clone is from Kohara lambda miniset library"
complement(1. .705)
/gene="rect"
complement(<1. .705)
/gene="rect"
/note="ORF_ID:0262#8
similar to [SwissProt Accession Number P33228]"
/codon_start=1
/transl_table=11
/product="Rect protein (P33)."
/protein_id="BAA14951.1"
/db_xref="GI:1742218"
/translation="MKQPPIAKADLOKTOGNRAPAAVKNSDVISFINOPSMKEQLAAALPRHMTAERMIRIATTEIRKVPALGNCDDTMSFSAIVQCSQLGLEPGSALGHAYLLPFGNKNEKSGKNVQLIGYRGMIDLARSGQIASARVVRGDEFSFEGDLEKLIH RGENEDAPVTHVYAVARLKDGGTQFEVMTRKQILEVRLSKAGNNGPWTHWEMAK KTAIRLLEKYLPSVSIET"
complement(698. .3298)
/gene="recE"
complement(698. .3298)
/gene="recE"
/note="ORF_ID:0263#1
similar to [SwissProt Accession Number P15032]"
/codon_start=1
/transl_table=11
/product="Exodeoxyribonuclease VIII (EC 3.1.11.-) (Exo VIII)."
/protein_id="BAA14952.1"
/db_xref="GI:1742219"
/translation="MSTKPLFLLRKAKKSSGEPDYVLMASNDFESTCATILDVLVKS GKLSYFKAVATNFPVYVNDLPAGEIDFTWSERYQSKDSMTWELKPGAAPDPAHYQG NTNNGEDMTEIEENMLPIISGQELPIRWLAQHGSEKPVTHVSRDGLQALHARAEE LPAVTALVSHKTSILDPLEIRELHLKLVRDYTDKVPENPNSNLGLITAFEEAYLNADYT DRGLLTKEWMKGNRVSHITRTASGANAGGNTDRGEGFVHDLTSLARDVATGILARS MDLDIYNLHPAHAKRIEETIAENKPPESVFRDKFTMPEGGLDYSRAIVASVKEAPIG IEVIPAHTVEYLNKVLFTEDHANPDPEIVDIACGRSSAPMPQRTVEGKQDDEKQPQ SGTVAEOGEAETMEPDATTEHHQDTQPLDAQSOVANSVDAKYQELRAELHEARKNIPSK NPVDDDKLIAASRGFEVDGISDPNDPKWVGIOTRDCVYQONQPEKTEKSPDMNPPEY VOQPEIACNAGGOTGDCNCPDCAVMGDATYQCFDEESOVEAKENDEEMGAEPH HNEAGSDPHRDCSDTEGEVADPVIVEDIIEPGIYGISNENYHAGPGISKSQLDIAD TPALYIMRKNAIPVDYTKTKTLDLGTAFCRVLEPEFSNREIVAPENRRTNAGKEE KAFLMCEASTGKTVTITAEGRKIELMYQSMALPIGOMLVESAGHAESSIYWEDEPETG ILRCRCPDKIIEPFHMTDYKTTADIQRKTAAYDYRVYHODAFYSDGYEAQFGVQPT FVFLVASTTIECGRYPVEIFEMGEEAKLAGQOEYHKNLRTLSDCLENTDEWPAIKTLSL PRAKEYAND"
complement(3400. .3675)
/gene="racc"
complement(3400. .3675)
/gene="racc"
/note="ORF_ID:0263#2
similar to [PIR Accession Number JS0119]"
/codon_start=1
/transl_table=11
/product="Racc protein"
/protein_id="BAA14953.1"
/db_xref="GI:1742220"
/translation="MITNYEATVVTDDIVHEVNEEGKRIGYVIKTENKETPFTVVDI DGPSGNVKTLDGCVKMKCLVHIGKNLPAEKKAEPFLATLIAKMKGEI"
complement(3750. .3920)
/gene="ydaE"
complement(3750. .3920)
/gene="ydaE"
/note="ORF_ID:0263#3
similar to [SwissProt Accession Number P38394]"
/codon_start=1
/transl_table=11
/protein_id="BAA14954.1"
/db_xref="GI:1742221"

FEATURES

source


```

/translation="MTKKIKCAYHLCKKDVEESKAIERMLHFHMGILSKDEPRKCYSE
ACAERKQMAHEL"
complement(3920. .4141)
/gene="ydaD"
complement(3920. .4141)
/note="ORF_ID:0263#4
similar to [SwissProt Accession Number P38393]"
/codon_start=1
/translation_table=11
/protein_id="BAA14955.1"
/db_xref="GI:1742222"
/translation="MIAHFGTDEIPROCVTPGDVYLHEGRTYIASANNIKKKLYIR
LNTTKFTIDRMIKVFLGRDGLPVKAESW"
4583. .5071
/gene="sleB"
4583. .5071
/note="ORF_ID:0263#5
similar to [SwissProt Accession Number P38392]"
/codon_start=1
/translation_table=11
/product="Superinfection exclusion protein B."
/protein_id="BAA14956.1"
/db_xref="GI:1742223"
/translation="MLDVFTPLKLFANEPLERLMYIIIFGLTLWLIPKEFTVAFNA
YTEIPWLFOIIVFAEFSEVVAISFSRLRAHIQKHYSLLPEQRYLLRSEKIAVEKDFL
KTGNLIITSPCRNPVMKKLERKGIIOHSDSANCSYLVTEKYSHFMKLEWNSRRF
NR"
complement(5068. .5223)
/gene="ydaF"
complement(5068. .5223)
/gene="ydaF"
/note="ORF_ID:0263#7
similar to [SwissProt Accession Number P38395]"
/codon_start=1
/translation_table=11
/protein_id="BAA14957.1"
/db_xref="GI:1742224"
/translation="MQKIDLGNNESLVCGVFPNQDGTFTAMTYTKSKTPEETGARRW
LEKHTVS"
7895. .8641
/note="ORF_ID:0263#10
similar to [SwissProt Accession Number P07905]"
/codon_start=1
/translation_table=11
/product="DNA replication protein DnaC."
/protein_id="BAA14958.1"
/db_xref="GI:1742225"
/translation="MKNIATGDVLERIRLAPSHVTAPFKTVAEWREMQLSEGOKRCE
EINRONROLREYEKILNRSGIQLPHRKCSFSNYQVONEGORALSOAKSIADELMTGCT
NEAFSGKPGTGKNHLAAAIQNRLLKDGQYIVTVADVMASALHASYDGSQSGEKLRE
LCEVDLLVDEIGIQIRETKNEQVVLHQIVDRRTASMRSVGMLTINLYEAMKTLGERI
MDRMTMNGGRMVNFWNEMSWRPNVQPGIAK"
9251. .9556

CDS

Alignment Scores:
Pred. No.:      72.1      Length:      18700
Score:          155.50     Matches:      121
Percent Similarity: 35.49%  Conservative: 60
Best Local Similarity: 23.73% Mismatches:    220
Query Match:      6.45%     Indels:      109
DB:              1        Gaps:      21

US-09-825-414-7 (1-486) x D90774 (1-18700)

QY  19 ArgThrAlaSerAspAlaSerLeuAlaSerSerValArgSerValSerSerAspGln 38
    ||| ||||| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::|
Db 15303 CGTAAACGGCTCGCGGTGGCACAGAACACGGCAGCCGGAAGTACAGCCAGTGATGCC 15362

QY  39 GlnArgGluIleAsnAlaIleAlaAspTyrLeuThrAspHisValPheAlaAlaHisLys 58
    ||| ||| ::| ::| ::| ||| |||||
Db 15363 AGCATATCAGCCCGGTGAGGGCGCAACCCATGCGGCTGAT-----GCTGGCGACTCA 15413
```

```

QY  59 LeuProProAlaAspSerAlaAspGlyGlnAlaAlaValAspValHisAsnAlaGlnIle 78
    ||| ::| ::| ::| ::| ::| ::| ::| ::| ::| ::|
Db 15414 GCACGGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 15473

QY  79 ThrAlaLeuIleGluThrArgAlaSerArgLeuHisPheGluGlyGluThrProAlaThr 98
    ::| ::| ::| ::| ::| ::| ::| ::| ::| ::|
Db 15474 AGCGCA-----GGAAGGCATCAACAAAGGCCACT 15503

QY  99 IleAlaAsp--ThrPheAlaLysAlaGluLysLeuAspArgLeuAlaThrThrThrSer 117
    ||| ::| ||| ||||| ||| ||| ::| ::| ::|
Db 15504 GAAGCATCAAAAAGTGCTGCCGTGCAGAGTCTCAAAAAGCGCGGCACACAGTGCC 15563

QY  118 GlyAlaLeuArgAlaThrProPheAlaMetAlaSerLeuLeuGlnTyrMetGlnProAla 137
    ||||| ::| ::| ::| ::| ::| ::| ::| ::|
Db 15564 GGTGGCGCGAAACGTCAGAAACGAATGCTTCAGCGTCACTACAAATCAGCAGCACATCT 15623

QY  138 IleAsnLysGlyAspTyrLeuProAlaProLeuLysProLeuThrProLeuIleSergly 157
    ::| ::| ::|
Db 15624 GCATCCACCGCG-----ACCACGAAGGCATCAGAA 15653

QY  158 AlaLeuSerglyAlaMetAspGlnValGlyThrLysMetMetAspArgAlaThrGlyAsp 177
    ||| ::| ||| ||||| ::| ::| ::| ::| ::|
Db 15654 GCTGGCAGCTCGGCCCGGATGGCGCGGCTCAAAAGAGCGCAAAATCATCA--GAA 15710

QY  178 LeuHisTyrLeuSerAlaSerProAspArgLeuHisAspAlaMetAlaAla----Ser 195
    ::| ::| ::| ::| ::| ::| ::| ::|
Db 15711 ACGAACGATCATCAAGCGCCAGTAGTGACGCTTCGCGCAACGCGGCAAGAAATTC 15770

QY  196 ValLys-----ArgHisSerProSerLeuAlaArgGln 206
    |||
Db 15771 GCGAAGCGCGCAAAAACGTCCGAGACGAACGCCAGGTCTTGAACGGCAGCGGACAG 15830

QY  207 ValLeuAspThrGlyValAlaValGlnThr-----TyrSerAlaArgAsnAlaValArg 224
    ||| ::| ::| ::| ::| ::| ::| ::|
Db 15831 AGCGCTCGGCTGCGGCGAGGCTCAAAACAGCGGCTGCGTCTGCCAGTGCAAGCGTCA 15890

QY  225 ThrValLeuAlaProAlaLeuAlaSerArgProAlaValGlnGlyAlaValAspLeuGly 244
    ||| ||| ||||| ||| ::| ::| ::|
Db 15891 ACAAGTCCGGGCGAGGCGCTCAGCCAGTGCCACCGCGCGGCAAAATCGGCAGAAAGCGCC 15950

QY  245 ValSerMetAlaGlyGlyLeuAlaAlaAsnAlaGlyPheGlyAsnArgLeuLeuSerVal 264
    ||| ||| |||||
Db 15951 GCATCGTCTGCTTCACACAGCCACACGAAGGCTGGGAAGCCACTGAACAGGCCAGCGCA 16010

QY  265 GlnSerArgAspHisGlnArgGlyGlyAlaLeuValLeuGlyLeuLysAspLysGluPro 284
    ::| ::| ::|
Db 16011 GCAGCGAGGTCT-----GCTTCGCGAGCGAAGACATCCGAACG 16049

QY  285 LysAlaGlnLeuSergluGluAsnAspTyrLeuGluAlaTyrLysAlaIleLysSerAla 304
    ||| ::| ||||| ::| ::| ::| ::| ::|
Db 16050 AACGCGAAGCGTCCGGAACAAGC-----GCAGAATCCTCAAAACGGCTGCCGCATCG 16103

QY  305 SerTyrSerglyAlaAlaLeuAsnAlaGlyLysArgMetAlaGlyLeuProLeuAspMet 324
    ||| ||| ||||| ::| ::| |||
Db 16104 TCAGCCAGTTCGGCGCGCTCATCGCATCATCG-----GCGTCTGCTCAAAAAGATGAG 16157

QY  325 AlaThrAspAlaMetGlyAlaValArgSerLeuValSerAlaSerSerLeuThrGlnAsn 344
    ||||| ||| ::| ::| ||| ||||| ::| ::|
Db 16158 GCGACCAAGACAGCGTCAGCAGCGAAGAGC-----AGCGCACAGCGGATCCACGAAG 16211

QY  345 GlyLeuAlaLeuAlaGlyGlyPheAlaGlyValGlyLysLeuGlnGluMetAlaThrLys 364
    |||||
Db 16212 GCGACAGAGGCTGTCGACGT----- 16232

QY  365 AsnIleThrAspProAlaThrLysAlaAlaValSerglnLeuThrAsnLeuAlaGlySer 384
    ||||| ||||| ||||| ||| ::| |||
Db 16233 -----GCGACGGCGGCAAGCTCAGAGCAAAAGTACG-----GCGGAATCC 16271

QY  385 AlaAlaValPheAlaGlyTyrThrThrAlaAlaLeuThrThrAspProAlaValLysLys 404
    ||||| ||| ::| ::| ::| ::| ::| ::|
Db 16272 GCGGCAACGGCGCGC-----GAGACAGCAGCTAAACGG 16304
```

QY	405	AlaGluSerPhe-----IleGlnAsp-----ThrValLys-----	414
Db	16305	GCAGAGGATATTGCATCCGCCCTGGCGCTTGAGGATGCCAAGTACGACGAAAAAGGGGATA	16364
QY	415	-----SerThrAlaSerSerThrThrGlyTyrValAlaAspGlnThrValLysLeu	431
Db	16365	GTACAGCTCAGCAGTGGCACCACAGTACGCTCTGAAAACGCTGGCGGCAACGCCAAGGCA	16424
QY	432	AlaLysThrVal-----LysAspMetGlyGlyGluAla	442
Db	16425	GTAATAATCAGCCTATGACAAATGCAGAGAAACGTCTGCAGAAAGACCAGAACGGCGTGAT	16484
QY	443	IleThrHisThrGlyAlaSerLeuArgAsnThrValAsnAsnLeuArg-----	458
Db	16485	ATACCCGATTAAGGATGCTTCCTG---AACACATTTAACCGGCTCAGTAAACAGACTTT	16541
QY	459	-----GlnArgProAlaArgGluAlaAspIleGluGluGlyGlyThrAlaAlaSerPro	476
Db	16542	GCTGATAAGCGGTGATATGCCGTTATGCGGGTTAACGCTCCTGCAGGTGCAACATCTTGA	16601
QY	477	SerGluIleProPheArgPrometArgSer	486
Db	16602	AAATATTACCCCTGTGTGTTATGCGTTCT	16631

RESULT	13
AF317649	
LOCUS	5329 bp DNA linear BCT 10-NOV-2000
DEFINITION	Rhodobacter sphaeroides flagellar hook-associated protein 1 (flgK) and flagellar hook-associated protein 3 (flgL) genes, complete cds.
ACCESSION	AF317649
VERSION	AF317649.1 GI:11136618

SOURCE ORGANISM	ORGANISM
Rhodobacter sphaeroides.	Rhodobacter sphaeroides

Rhodobacter.

REFERENCE 1 (bases 1 to 5329)
AUTHORS Gonzalez-Pedrajo, B., De La Mora, J., Ballado, T., Camarena, L. and

TITLE	Isolation and Complementation of a Flagellar P-Ring Mutant of Rhodobacter sphaeroides
JOURNAL	Unpublished

JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 5329)
AUTHORS	Gonzalez-Pedrajo, B., De La Mora, J., Ballado, T., Camarena, L. and Dreyfus, G.

TITLE Direct Submission
JOURNAL Submitted (31-OCT-2000) *Genetica Molecular, Instituto de Fisiologia Celular, UNAM, Circuito Exterior, Ciudad Universitaria, Mexico*
Distrito Federal 04510, Mexico

FEATURES
source

```

/organism="Rhodobacter sphaeroides"
/db_xref="taxon:1063"

```

gene
CDS

```

/gene="flgK"
/note="HAP 1"
/codon_start=1
/transl_table=11
/product="flagellar hook-associated protein 1"
/protein_id="AAG31286.1"
/db_xref="GI:11136619"
/translation="MSILDIARSGILSYRSALSVTAENIANVNTTEGYVREAVLAQVP
GGQISPTSAATSGQGVREVDVRRAFDGLVAQRKRTAEGAAASAETLDTTGALESFL
AGAGSVPEALITGEFDALNGLGATPRDGLRQVLAAGQTLIASFYTAASGLAGLDEYV
TGLAGQTAVEASQTLRQLAELNGRMTGAQOASALNPLLEDERDLTGELSRFTVGVOASFD
ALGRFTYTLGSAPGPLLLEGTAASTLGVAAEAGGLVLEVRTAGATLQSRQITGGSLOG
YASALSAVDNAAAEELDALARQLVAENMTVHAQGIDQSGARGDILFALEGWQVAAADAGN
RGTTGASVVAEFDATATAPGLITLHDAAGLWHAYDAGTLLGSGAETLALTGLTQMT
GAAMDGDRLTVSPRQGRALDMRFLITDTROIAAAPYIVSAASGNAGSAQASFVPTVS
TLPATGSLGLLTDPAGADATLIRAGVYGHIPAGTSVTLASLGSQSQQEFLLSDPA
AAGATLTSFSIGGTIVHSFDLSALGATDPATLAAALNAGAVSTGGQTLDELGVTAGSST

```

gene

CDS

4100..5329

/gene="figL"

4100..5329

/gene="figL"

/note="HAP 3"

/codon_start=1

/transl_table=11

/product="flagellar hook-associated protein 3"

/protein_id="AAG31287.1"

/db_xref="GI:11136620"

/translation="MTLGTTLFATLASRNFSLQTEIGTLQARIASEVRDPRPSADPT
RAVQLSAKEMEATLGRYGANASSADRLAHADVALGEVAARMEKDVVLQAGNPTL
SDEGRAGLRIVAESAREALFALANRDAMQGLFAGYAAGPAVKEGDTVRFAGNDGO
PVAQLSETLVRVATSLGGAEVEMAVPEAGPRSVFELADDLVATLTVPLAHSSPQRTAE
DTARLSLAAPPAPATVRFRTLTPGGSVEIEQRLPGSALAAINASPQGTATEEPDG
TLVLGAVGRITVSDMSRSDDPRIYALFTSADDKGGWIMPARDAASLTDAFDAVSH
MAEORARAGALAAASVDTOADAIRTRQTRIATAIGGLEDDLVAAEAVTRLQQLLLTQEA
QQTIVYKIANRSLFDYLR"

Alignment Scores:	
Pred. NO.:	24
Score:	151.50
Percent Similarity:	36.77%
Best Local Similarity:	27.10%
Query Match:	6.28%
DB:	1
Length:	5329
Matches:	126
Conservative:	45
Mismatches:	177
Indels:	118
Gaps:	24

US-09-825-414-7 (1-486) x AF317649 (1-5329)

QY	85	ArgAlaSerArgLeuHisPheGluGlyGluThrProAlaThrIleAlaAspThrPheAla	104
Db	1956	CGGCTTTCGGGAAACCGGGAATCGGGCGGATCCTGCCGGGCGAGCGTGCTGTGGCT	20155
QY	105	LysAlaGluLysLeuAspArgLeuAlaThrThrThrSerGlyAlaLeuArgAlaThrPro	124
		::: ::: :::	
Db	2016	TCATCCCGCGCA-----TCCGGTGGCCGGCAGTTCCGGGCT	2051
QY	125	PheAlaMetAlaSerLeuLeuGlnTyrMetGlnProAlaIleAsnLysGly-AspTyrPhe	144
		::::: ::	
Db	2052	-----TCTGCCCTCGCATCCGCCCTCCCGAGCCTGAGCGTGAAGCGCGCGGGCCT	21055
QY	144	uProAlaProLeuLysProLeuThrProLeuIleSerGlyAlaLeuSerGlyAlaMetAs	164
		::: ::: :: :::	
Db	2106	TCCCTCCCGGTGCAG-----CTGCCGTGGGGTGCCTCGGGCGCGAGATGGC	21535
QY	164	pGlnValGlyThrLysMetMetAspArgAlaThrGlyAspLeuHisTyrLeuSerAlaSe	184
		:::::	
Db	2154	GCAGGCCAT-CAACGCGTTCCGGGCGCGGATCGAGC-----CGAGG	2194
QY	184	rProAspArgLeuHisAspAlaMetAla-----AlaSerValLys-----Ar	198
		:::	
Db	2195	CCCGCACCGGCGGTACAGATCGAGGGCGCGCCGACGGCAGCCTACGTTCCGCGCTACCG	2254
QY	198	gHisSerProSerLeuAlaArg-GlnVal-----LeuAspT	210
		::: :: ::	
Db	2255	GCACCAACT--CTGCGCCGTGGCATCAGCGGGCGCGCGGGGGGGCGCATGATG	2311

Qy 210 hr---GlyValAlaValGlnThrTyrSerAlaArgAsnAlaValArgThrValLeuAlaP 229
Db 2312 CGCTGGCGCTGGCGGTCAATGCGTGACCGGCCACCAGCGGTGCGGCCGAGCTGTGCG 2371
Qy 229 roAlaLeuAlaSerArgProAlaValGln---GlyAlaValAspLeuGlyValSerMeta 248
Db 2372 CGGACGGGGCGCGCTCCTCCTCGTTCAGAGCGGGGGCGGATATCGGATCGTG---- 2427
Qy 248 laGlyGlyLeuAlaAlaAsnAlaGlyPheGlyAsnArgLeuLeuSerValGlnSerArga 268
Db 2428 -----GGCTTCGCCACACACGCGCGGGCGCGCTCACGCTGCAGGGCACGACGCCGAGG 2482
Qy 268 spHisGlnArgGlyGlyAlaLeuValLeu---GlyLeuLysAspLysGluProLysAlaG 287
Db 2483 GCAGTCCCGCAGCGCGCCCTCACCCTTGTCGGGCACGGCCGAC-----AGCGCGC 2533
Qy 287 lnLeuSerGluGluAsnAspTrpLeuGluAlaTyrLysAlaIleLysSerAlaSerTyrS 307
Db 2534 GCTTCACGGCGCAG-----CTGCGGCTCTCGTCGCGCA 2566
Qy 307 erGlyAlaAlaLeuAsnAlaGlyLysArgMetaIaGlyLeuProLeuAspMetaIaThra 327
Db 2567 GCGGTTTCGCGCGCATCTGGC-----GGCGTCGGCAGGATGCGCGCGTGG 2614
Qy 327 spAlaMet-----GlyAlaValArgSerLeuValSerAlaSerSerLeuThrGlnAsn- 344
Db 2615 ATCCGATGAGCGCGCGCTCTCGCGCGCGGTGAGCGGGGACGGCGGTTCAGACCT 2674
Qy 345 -----GlyLeuAlaLeuAlaGlyGlyPhea 353
Db 2675 ATGGCTTACCTTACGATCCCGCCTTCAGCGGGCGGGGCTCTCGCGGACGCGACCTTCG 2734
Qy 353 laGlyValGlyLysLeuGlnGluMetaIaThr-----LysAsnIleThrAspProA 370
Db 2735 CGCAGGCGGTTCCGCGCAGTATGCGATGACGTGGGCCAACCGGACGCTGACACTCGATG 2794
Qy 370 laThrLysAlaAlaValSerGlnLeuThrAsnLeuAlaGlySer----- 384
Db 2795 CCGCGGCGCGCGCTGAGCAGCGGGCGCGCTGCGCTTCGCGCTTCCTCTGC 2854
Qy 385 -----AlaAlaValPheaIaGlyTrpThrAlaAlaLeuThrAsp 400
Db 2855 GCGCCGAAGCCCCCGCGCGCAGCGTACCGCGCGGCCCGCTCGCGCGCTG-----C 2905
Qy 400 roAlaValLysLysAlaGluSerPheIleGlnAspThrValLysSerThrAlaSerSer 420
Db 2906 CCGCCGATGCGCGCGCTCGTCTCCTTACGAGGGGACAGCTACACGCTGCCCATGA 2965
Qy 420 hrThrGlyTyrValAla-----AspG 427
Db 2966 CGGGTGGGGCGGTCCGGTGCATGGCGCGGACCGCGGACTGCTCACGGCGCGCTTCGACG 3025
Qy 427 lnThr-----ValLysLeuAlaLysThrValLysAspMetGly----- 439
Db 3026 CGACGAACCGTCTGTCATCCAGCGCGGCAAGTTGCTCGACGGCGCGGCTGCGCATCG 3085
Qy 440 -----GlyGluAlaIleThrHisThrGlyAlaSerLeuArgAsnThrValAsnLeuA 458
Db 3086 AGAGCGGGGCGCGCGCGCTTCGAGCTGGCGCGCGCGCGCGCATCA 3145
Qy 458 rgGlnArgProAlaArgGluAla-----AspIleGluGluGlyGlyT 472
Db 3146 CCGGTCAGCCCGCGGATCCCGCGCGCTGCCCGCGAGCTTCGACATGAGCTGGGCGGCA 3205
Qy 472 hrAlaAlaSer 475
Db 3206 CCCTCTACAGC 3216

RESULT 14
AL646078 203050 bp DNA linear BCT 07-DEC-2001
LOCUS Ralstonia solanacearum GM11000 megaplasmid, complete sequence;
DEFINITION segment 3/11.

ACCESSION AL646078 AL646053
VERSION AL646078.1 GI:17430778
KEYWORDS
SOURCE Ralstonia solanacearum.
ORGANISM Ralstonia solanacearum
Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
Ralstonia.
REFERENCE 1 (bases 1 to 203050)
AUTHORS Salanoubat,M., Genin,S., Artiguenave,F., Guzy,J., Mangenot,S.,
Arlat,M., Billault,A., Brotlier,P., Camus,J.C., Cattolico,L.,
Chandler,M., Choisine,N., Claudel-Renard,C., Cunnac,S., Demange,N.,
Gaspin,C., Lavie,M., Moisan,A., Robert,C., Saurin,W., Schlex,T.,
Siglier,P., Thebault,P., Whalen,M., Wincker,P., Levy,M.,
Weissenbach,J. and Boucher,C.A.
TITLE Genome sequence of the plant pathogen Ralstonia solanacearum
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 203050)
AUTHORS Boucher,C.A.
TITLE Direct Submission
JOURNAL Submitted (05-DEC-2001) Genoscope and CNRS UMR-8030, 2 rue Gaston
Cremieux, CP5706, 91057 Evry Cedex, France, Laboratoire de Biologie
Moleculaire des Interactions Plantes-Microorganismes INRA-CNRS,
BP27, 31326 Castanet-Tolosan Cedex, France, Fondation Jean
Dausset-CEPH, 27 rue Juliette Dodu, 75010 Paris, France, LMGM CNRS
118 Route de Narbonne, F 31062 Toulouse Cedex 4, Genoscope and INRA
URGV, 2 rue Gaston Cremieux, CP5706, 91057 Evry Cedex, France,
Laboratoire de Biometrie et Intelligence Artificielle INRA, BP27,
F31326 Castanet-Tolosan Cedex, Laboratoire de Genetique Cellulaire
INRA, BP27, F31326 Castanet-Tolosan Cedex
Christian.Boucher@toulouse.inra.fr
COMMENT http://sequence.louluouise.inra.fr/R.solanacearum.html.
FEATURES
source location/Qualifiers
1. .203050
/organism="Ralstonia solanacearum"
/strain="GM11000"
/db_xref="taxon:305"
/plasmid="megaplasmid"
complement(147. .1499)
/gene="RSp0310"
/note="RS05460"
complement(147. .1499)
/gene="RSp0310"
/function="cell processes; transport of small molecules"
/note="Product confidence : probable
Gene name confidence : hypothetical
predicted by Codon_usage
predicted by Homology
predicted by Framed"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="PROBABLE MACROLIDE-EFFLUX TRANSMEMBRANE PROTEIN"
/protein_id="CAD17461.1"
/db_xref="GI:17430779"
/translation="MNAPRKNTDRAEAALLRNASTTHEPANPCWIPDAAAAGLHGP
WRARFWAIFGGQALSLIGSALTQFVLLWITDTGVSALATAGLALLPQAVLSPLG
GTFADRYSRRLMIVADGISALMLLIALFLTGRIALWHAIVMAIRSAQAFOABA
AAASVTMLVPGGFLVRAAGLNSLQSLTVVAAPLGAALGVMPIGWALGIDVATALL
GIAPLICRIPQACVSGGKGLMPAFREGVDLVKTPGLRWLVVLGAVLITIMPTF
TLVPLLVKTHFGGAAQVLALEGLSIGVAGGLVYANAPRRQVAMWILCGFANSCVA
IALTALPGLSFGVAVAMWVTSITVEFGNAPLTALQTIVPNHLOGRVLSLNTMG
LAAPVGLALFTPLGEVIGVRAIFAIALGLGAASLAGFLSPSLMRDRPSRSEARTD
"

gene complement(1531. .2961)
/gene="oprM"
/note="RSp0311; RS05459"
complement(1531. .2961)
/gene="oprM"
/function="cell processes; transport of small molecules"
/note="Product confidence : probable
Gene name confidence : putative
predicted by Codon_usage
predicted by Homology

CDS

predicted by Framed"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="PROBABLE OUTER MEMBRANE CHANEL LIPOPROTEIN"
/protein_id="CAD17462.1"
/db_xref="GI:17430780"
/translation="MRLRLPLVLALSACSLPPALIKPALPVPATFPAESPEETOAHA
ADLGWRTMGDRRLQRLIGLTLANNRDLRLALNVEITVRAQYSIQRAARLPGIDATGS
ASRQRTANDGASPALAANSEQYGVNGLNAFEIDLFGVRVSLSDAARLYASDQG
RRAQISLVGADAVYFAERLAEQORQLAEHTLADWRQSLDLARLKTAKONSDLVA
QAEQVANAGADIEARTALAAQAGNSLQRLVGAGLPADLPAPLSLETQPVVTLQPLAGL
PSELIRRPDIRGAENHLAANADIGAARAFPPRLSLTASIGYASPAAGSLFDGHR
VWSFSPQITQPLEQGGRLRAELRLAELRKSSALAEYERTIQTAFREVADGLGRATYG
ROIENQIRAVASERVELSNLRYRAGVEGRLELDSQRLYASRQALDLRQGAFSN
AVALYKALGGIAETDVATADRVARK"
complement(2966..6115)
/gene="mexD"
/note="RSp0312; RS05457"
complement(2966..6115)
/gene="mexD"
/function="cell processes; transport of small molecules"
/note="Product confidence : probable
Gene name confidence : putative
predicted by Codon_usage
predicted by Homology
predicted by Framed"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="PROBABLE MULTIDRUG EFFLUX SYSTEM TRANSMEMBRANE
PROTEIN"
/protein_id="CAD17463.1"
/db_xref="GI:17430781"
/translation="MPQEFIRRPVFAWVIALFTILLGAIAIPKLPRIARVPSVAPPTVN
LSATYPGATPQTMNDGVISLIERLSVKNLTFESSSDTSGSASIVATFKPGTDPM
AQVDQNKLKAIERLPQVVRONGLTVESASGFLMIIGLQSDGREFDETALSDYMAR
NVEELRRIEGVRLVQFGAEQAQRIWDPAKLIAYGLSMNDLSAISQNAQIAPGS
VGASPALPAQRTVYPLTVQGQLTPEQFAIYLRANANGSKVMGDVARELGAQSYG
FINRENGKAATAAVQLAPGANVRTAEGVOARMAELSRTPAGMSYSIPNTAPEVK
ISIEKVIQTLIEAMVLFVLMYLFQNVRYTLIPAIAPIALLGFTTILASGSFINV
LTMFGMVLAIQIVDDAIVVENVERLMAOEGLSPRDATIKAKEITGAVIGITLVLT
AVFIPMALASGSYGVYIQQFTLSMAVSIVFSAPLALTLPALCATLLKPYDPAHHEKR
GVFGFNRRFRDRTQGYAARVLTGRTGRVMEFGAIAAVLALAFRQLPSSFLBEED
QGYEMTFQLPADATAERTLDVYKAFERHVASRASISQNSMISIVGFGSGSNAAMAF
TMLKWDARNGATAQONEMRLAQAMAGATEGTVMSLLPPAIDELGTSSTGFTLRLQDRA
NQGDAALKAETKLELAAQSRIVSQVYPDGLPAGTSIRLDIDRQKAEALGVAFTSIS
ETLSTAIGSLYVNDFPNAGRMQOVIQADAPARMOIDDVLYVRNAGGCVPLSEVY
RPVWSETPLQLVRFQGFPAARISGNAPGASSGAMAEMERLAAQLPGFAIAMTGOS
LOEROSAAQAPMLMALSLMVLVFLVALALYESNAIPLSVMLVPLGLLGAVALLRGM
PNDVFEKVMITIVIGLSAKNAIILIVEFAKOLREOGKGLVESAVTAAARLRPLIMTSL
AFGLGVPLMIAMGASAEQTHAIGTGVFGGVSATVLAIFVFVFVFMGLQERFEA
WRASRQAPPPVPHPQPEG"
complement(6136..7293)
/gene="mexC"
/note="RSp0313; RS05456"
complement(6136..7293)
/gene="mexC"
/function="cell processes; protection responses;
drug/analog sensitivity"
/note="Product confidence : probable
Gene name confidence : putative
predicted by Codon_usage
predicted by Homology
predicted by Framed"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="PROBABLE MULTIDRUG EFFLUX SYSTEM PROTEIN"
/protein_id="CAD17464.1"
/db_xref="GI:17430782"
/translation="MKIERLRHGGVHLLAAMLACGCPSEPOSAPPAAEVATTTVAP
AHLELTEDLPGRVAAVRAVEIRPVQVSGIVQRRLFEGQTEVRAGQPLFQINAPFRADA

DTAAALRRAEALALARAVQTRLQPLVEADAVSRQAYDDAVSQRDQAADVEQARAT
LAROLDLKFATVEAPIGRIDQALVEGALVSSGDSNPMARIQIDQVYDVRRPVS
SLEALRQTLATQKTGAGDGLPVAVLRSNGEPYEAKGHILFSGINVDAGTDVLLRVLS
NMPORQLPGMEVVARVPHARYTDLTVPOQAVVRGPQVWTFIDANSHARLKPVEL
GELADRRYRIRAGLSAGOKIVEGMERLSDGAAYAAHDKMSPEPVAGPAR"
7392..8162
/gene="RSp0314"
/note="RS05454"
7392..8162
/gene="RSp0314"
/function="miscellaneous; not classified regulator"
/note="Product confidence : probable
Gene name confidence : hypothetical
predicted by Codon_usage
predicted by Homology
predicted by Framed"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="PROBABLE TWO COMPONENT RESPONSE REGULATOR
TRANSCRIPTION REGULATOR PROTEIN"
/protein_id="CAD17465.1"
/db_xref="GI:17430783"
/translation="MEILNORGAAAVPMPBPSCGAOSAQALILIAEDEPEIADILAA
YLARSGLRSVHADGRLELHLSLKPDVLVLDVQMPHYDQWKAALSEIRHRGDTPIVIM
LTALDQDIDKLMGRIGADDYVVKPFPNPAEYVARQAVLRSMSCGTRDQRLRVAP
FQIDLDNHEATVQYQKQKRALALTTEFKLILQALARAPKRVFSRAEILLAACLEPBDTL
ERTVDSHISKLRKLEDLGVNGMPGVGRGYRLWSLE"
8159..9241
/gene="RSp0315"
/note="RS05453"
8159..9241
/gene="RSp0315"
/EC_number="2.7.-.-"
/function="miscellaneous; not classified regulator"
/note="Product confidence : probable
Gene name confidence : hypothetical
predicted by Codon_usage
predicted by Homology
predicted by Framed"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="PROBABLE TRANSMEMBRANE SENSOR KINASE
TRANSCRIPTION REGULATOR PROTEIN"
/protein_id="CAD17466.1"
/db_xref="GI:17430784"
/translation="MKLAGLSRQIALSMMAAFGVTLLVLTSAFYIAMIYWSRDF
COASWLPPTGPEWMLIATTLIAGLAMAIVVNLARILVPLNSVADSLRVARGDLGV
RAIAGDRSLGEAALLADDFNALADQLQRYTEBOVFWNAAIAHELRTPTVTLRGRLOGL
AEGVFTPDEPQRSLLTQVESLTRLIEDLRVYSLAESGHLDLQVRESDLVAEVKAVVA

Alignment Scores:
Pred. No : 2.54e+03 Length: 203050
Score: 151.50 Matches: 125
Percent Similarity: 33.59% Conservative: 48
Best Local Similarity: 24.27% Mismatches: 213
Query Match: 6.28% Indels: 130
DB: 1 Gaps: 19

US-09-825-414-7 (1-486) x AL646078 (1-203050)

Qy 10 ProProvalThrAlaThrAspSerPheArgThrAlaSerAspAlaSerLeuAlaSerSer 29
||| ||||| ||| ||| |||||
Db 94689 CCCAATACCGCCGCGACCGACCTGGCGATGCTGTCGACGCGCGGACGCGCGGAC 94748
Qy 30 SerValArgSerValSerSerAspGlnGlnArgGluIleAsnAlaIleAlaAspTyrIleu 49
::: ||||| ||| :::: ||| |||
Db 94749 AAGTCGACAGCGCGCTGCTCGGCGCCGATACCTTCGGAACCTGCTGCGATCGCCTG 94808
Qy 50 ThrAspHisValPheAlaIleHisLysIleuProProAlaAspSerAlaAspGlyGlnAla 69
||||| ::| ||| ||||| ::|||
Db 94809 AACCAAGAG-----GCCGCGCGCGCCAGACCGATGCGCAATCGCGCGGACCGCGCC 94862

QY 70 AlaValAspValHisAsnAlaGlnIleThrAlaLeuIleGluThrArgAlaSerArgLeu 89
|||
Db 94863 GCC-----GCCCCCGC--- 94874
QY 90 HisPheGluGlyGluThrProAlaThrIleAlaAspThrPheAlaLysAlaGluLysLeu 109
Db 94875 -----GCCGATGCCAGGAAGAA 94892
QY 110 AspArgLeuAlaThrThrSerGlyAlaLeuArgAlaThrProPheAlaMetAlaSer 129
|||
Db 94893 GACCGCGCCAAAGCCAGCGCGCGCAAC----- 94922
QY 130 LeuLeuGlnTyrMetGlnProAlaIleAsnLysGlyAspTrpLeuProAlaProLeuLys 149
Db 94923 --GATCAGGACGCGCTCAGCGCCAGCCGCAATGCCGAGCGCGCGCGC----- 94970
QY 150 ProLeuThrProLeuIleSerGlyAlaLeuSerGlyAlaMetAspGlnValGlyThrLys 169
Db 94971 -----ACCGCGCGCGCTTCGGCCACGCGAGATGCACAGGCTTCAG 95012
QY 170 MetMetAspArgAlaThrGlyAspLeuHisTyrLeuSerAlaSerProAspArgLeuHis 189
Db 95013 -----GCCGCGCTCGAGCGAGCACGACGACACCAG 95036
QY 190 AspAlaMetAlaAlaSerValLysArgHisSerProSerLeuAlaArgGlnValLeuAsp 209
|||
Db 95037 GACAAGACCGCGCGCGCGCGCGCGCTCGATCCCGC----- 95075
QY 210 ThrGlyValAlaValGlnThrTyrSerAlaArgAsnAlaValArgThrValLeuAlaPro 229
Db 95076 GCACAGCTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTCCG 95135
QY 230 AlaLeuAlaSerArgProAlaValGlnGlyAlaVal-----AspLeuGlyValSerMet 247
Db 95136 -----GTGTCCAGCCCGGACGCAAGTCTCGAGCGCGCGAGATGCCGCGCAGCTC 95189
QY 248 AlaGlyGlyLeuAlaAlaAsnAlaGlyPheGlyAsnArgLeuLeuSerValGlnSerArg 267
Db 95190 GCGCGCGCGCAAGCG 95249
QY 268 AspHisGlnArgGlyGlyAlaLeuValLeuGlyLeuLysAspLysGluProLysAlaGln 287
Db 95250 GCCAATGCCGCGCGCGCAATGACC-----GATCCGGCGCGCCACC 95291
QY 288 LeuSerGluGluAsnAspTrpLeuGluAlaTyrLysAlaIleLysSerAlaSerTyrSer 307
Db 95292 GTCGCCACAGCG-----CAGGCTGCGCAAGCCGACGCCCTCAAG 95330
QY 308 GlyAlaAlaLeuAsnAlaGlyLysArgMetAlaGlyLeuProLeu--AspMetAlaThr 326
Db 95331 GCGCGCGCGCGCAAGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 95390
QY 327 AspAlaMetGly-----AlaValArg 333
Db 95391 GTCGCCACAGCGCTGCG 95450
QY 334 SerLeuValSerAlaSerSerLeuThrGln-----AsnGlyLeuAlaLeu 348
Db 95451 CACCTGGGCAAGCATCTGCTCCGCGCAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 95510
QY 349 AlaGlyGlyPheAlaGlyValGlyLysLeuGlnIleMetAlaThrLysAsnIleThrAsp 368
Db 95511 CACCGCACCGAAGCGAG 95570
QY 369 ProAlaThrLysAlaAlaValSerGlnLeuThrAsn-----LeuAlaGlySer 384
Db 95571 ACCGCGCGCAACGCGGACACGCGCGCAACCCCAACACAGCGCGCGCGCGCGCGCGCG 95630
QY 385 AlaAlaValPheAlaGlyTyrThrThrAlaAlaLeuThrThrAspProAlaValLysLys 404
Db 95631 GCGGCCAAGTTGATGGCTCTGACGCGCGCG-----AACGGGATGCCGTCCGCCACC 95684

QY 405 AlaGluSerPheIleGlnAspThrValLysSerThrAlaSerSerThrGlyTyrVal 424
Db 95685 GCGCCAACCTTC-----ACCGTGGCGCGCGCGCGGTACCGCGCAGCGCAGCGAGC 95735
QY 425 AlaAspGlnThrValLysLeuAlaLysThrValLysAspMetGlyGlyAlaIleThr 444
Db 95736 GCGCGGTCACACCCGCGCAGGGGCATACCTGCCGACCTTCGGCATGCCGCTGGCGCG 95795
QY 445 HisThrGlyAlaSerLeu-----ArgAsnThrValAsnAsnLeu 457
Db 95796 CACACCATGGCGAGCCAGCTGGCCTACATGCAGGTCCATCGCCAGTCTCCGCGCACTG 95855
QY 458 ArgGlnArgProAlaArg-----GluAlaAspIleGluGlyGly 471
Db 95856 CAGTTGAGCCCGCGCGAGCTCGCGCGGCTGCATGTCAAGCTTCGAGATCGACAACGGCGCT 95915
QY 472 ThrAlaAlaSerProSerGluIleProPheArgProMetArgSer 486
Db 95916 GTCAATGCCAGCT--TCGTCTGTACACACAGCGCGCGTGGCGGAAC 95958

RESULT 15
MTCY493 40790 bp DNA linear BCT 03-AUG-2001
LOCUS Mycobacterium tuberculosis H37Rv complete genome; segment 63/162.
DEFINITION 295844 AL123456
ACCESSION 295844.1 GI:3250713
VERSION
KEYWORDS
SOURCE
ORGANISM

REFERENCE
AUTHORS
Mycobacterium tuberculosis H37Rv.
Mycobacterium tuberculosis H37Rv
Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
Actinomycetales; Corynebacterineae; Mycobacteriaceae;
Mycobacterium; Mycobacterium tuberculosis complex.
1 (bases 1 to 40790)

REFERENCE
AUTHORS
Cole,S.T., Brosch,R., Parkhill,J., Garnier,T., Churcher,C.,
Harris,D., Gordon,S.V., Eiglmeier,K., Gas,S., Barry III,C.E.,
Tekala,F., Badcock,K., Basham,D., Brown,D., Chillingworth,T.,
Connor,R., Davies,R., Devlin,K., Feltwell,T., Gentles,S.,
Hamlin,N., Holroyd,S., Hornsby,T., Jagels,K., Krogh,A., McLean,J.,
Moule,S., Murphy,L., Oliver,S., Osborne,J., Quail,M.A.,
Rajandream,M.A., Rogers,J., Rutter,S., Seeger,K., Skelton,S.,
Squares,S., Squares,R., Sulston,J.E., Taylor,K., Whitehead,S. and
Barrell,B.G.
TITLE
Deciphering the biology of Mycobacterium tuberculosis from the
complete genome sequence
JOURNAL
Nature 393 (6685), 537-544 (1998)
MEDLINE
PUBMED
98295987
9634230

REFERENCE
AUTHORS
TITLE
JOURNAL
2 (bases 1 to 40790)
Parkhill,J.
Direct Submission
Submitted (11-JUN-1998) Submitted on behalf of the Mycobacterium
tuberculosis sequencing and mapping teams, Sanger Centre, Wellcome
Trust Genome Campus, Hinxton, Cambridge CB10 1SA Unite de Genetique
Moleculaire Bacterienne, Institut Pasteur, 28 rue du Docteur Roux,
75724 Paris Cedex 15, France E-mail: parkhill@sanger.ac.uk
On Jun 24, 1998 this sequence version replaced gi:2131042.

COMMENT

Notes:

Details of M. tuberculosis sequencing at the Sanger Centre are
available on the World Wide Web.
(URL, http://www.sanger.ac.uk/Projects/M_tuberculosis/) CDS have
been renumbered from the original cosmid submissions but the old
gene designations are in brackets after the new gene numbers.
Gene prediction was based on a Hidden Markov Model of TB genes
implemented in TBParse (Krogh) supplemented with visual inspection
of positional base preference in codons, especially where there is
an increase in the observed/expected third position G + C.
CAUTION: In some cases we may not have predicted the correct
initiation codon. Where possible we choose an initiation codon
(atg, gtg, or ttg) which is preceded by an upstream ribosome
binding site sequence (optimally 5-13bp before the initiation
codon). If this cannot be identified we choose the most upstream
initiation codon.

FEATURES

Location/Qualifiers

source 1. .40790
/organism="Mycobacterium tuberculosis H37Rv"
/strain="H37Rv"
/db_xref="taxon:83332"
<1. .1638
/organism="Mycobacterium tuberculosis H37Rv"
/strain="H37Rv"
/db_xref="taxon:83332"
/clone="Y21B4"
<1. .40722
/organism="Mycobacterium tuberculosis H37Rv"
/strain="H37Rv"
/db_xref="taxon:83332"
/clone="Y493"
complement(10. .771)
/gene="Rv1424c"
complement(10. .771)
/gene="Rv1424c"
/note="Rv1424c, (MTCY21B4.42c,MTCY493.30), len: 253.
Possible membrane protein, contains PS00402
Binding-protein-dependent transport systems inner membrane
comp signature."
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv1424c"
/protein_id="CAB09246.1"
/db_xref="GI:2131072"
/db_xref="SWISS-PROT:P71693"
/translation="MTVVGAPSRPASVSRPSYRQCVOASQTSARRYSEPPSYRRP
AEKLVPEYLLGILTLTLLSACQTASAGYNEPRGYDRATLKLVSMDGMLNRTYDS
KLAPSRPOVYACDSREARIRNDGFHANAPSCMRIDYELITQNHRAYYCLKLVRVGYC
YPAVTPGKPPSVLLYAPASACDESLSPRVATALVPGTRSANRESRFVVTEIKSLGA
GGRCDSASVSLQPEEIEGPAIRPASSQLVCVAPK"
complement(520. .606)
/gene="Rv1424c"
/note="PS00402 Binding-protein-dependent transport systems
inner membrane comp signature"
775. .2154
/gene="Rv1425"
775. .2154
/gene="Rv1425"
/note="Rv1425, (MTCY21B4.43,MTCY493.29c), len: 459.
Similar to many M. tuberculosis hypothetical proteins e.g.
MTCY28.26, fasta scores; E(): 0, 36.3% identlty in 460 aa
overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv1425"
/protein_id="CAB09245.1"
/db_xref="GI:2131071"
/db_xref="SWISS-PROT:P71694"
/translation="MKRLSSVDAAFWSAETAGWHMHVGALACDSDAPEYSFQRLRE
LIERLEIPQLRWRTGALPMELVIEGVEGRIATLTMMHAIYDGVSGGLGELLDTPEP
GRIMSYKIDRSRPLMELWVIEGVGGRIATLTMMHAIYDGVSGGLGELLDTPEP
RPPQETVGFVGFQIPGLERRAIGALINVGIMTPFRIVRLLEQYRQOIALGVAGKP
ARYFEAPKTRFNAPVSPHRRVTGTRVELARAKAVKDAFGVKLNDVVALVAGARQYL
QKRDELPAKPLIAQIPVSTRSEETKADVGNQVSSMTASLATHIEDPAKRLAIIHESLT
SAKEMAKAPSAHQIMGLTETTPGLIOLARAYTASGLSHNLAPINLVSNVPGPFP
LYMAGARLDSLVPPLGPPVMDVALNITCFSYQDYLDGLVTTPEVANDIDEMADAIEPA
LAELERAAE"
complement(2176. .3438)
/gene="lipo"
complement(2176. .3438)
/gene="lipo"
/note="Rv1426c, (MTCY493.28), len: 420. Unknown,
possibleesterase, similar to MTCY493.28, MTCY08D5.15,
MTCY21B4.16chighly similar esp. in C-terminus to M.
tuberculosis hypothetical protein Q50681 (431 aa) (50.1%
identity in 411 aa overlap), and also similar in central
region to AAAD_HUMAN P22760 arylacetamide deacetylase (398
aa), fasta scores; opt:210 z-score: 248.8 E(): 7.6e-07,
29.3% identity in 191 aaoverlap"
/codon_start=1

/transl_table=11
/product="lipo"
/protein_id="CAB09244.1"
/db_xref="GI:2131070"
/db_xref="SPTREMBL:006832"
/translation="MFRFRMARPRPLTRAVELLNANGLRPLSGSGYSTVLAFWLGM
PTSEVPGVYLGASVLDALRGRGDEGLKGKALALTAAMVILAVIRGATTPGP
VLEAGLTBQLGPDYAKELATLPTEPMRSRGRNLPRTAMARRRYETTNVCYGPYGR
ANLADIVRRRDLPDCAKAPLVLYVPGGAWLVGWRPQAYPLMSHLAARGWCVSLNRY
VSPRHTWEDHIVDKRALAWKENIAAYGGDPNFVAISGSGAGHLCALALTPNDR
FQPGFEQVDTSVAAAPVYGRYDFTTDAGRREFVGLTEFVVKRKFESTRDIFVDA
SPIHHVADAPPEFVLHGRHDSLIPVAEHAHAFVEELRAVSKSPVAYADLPHAQHAFDV
FGSPRAHHTAEAVARFLSWVYATNPAT"
complement(3438. .5045)
/gene="fadd12"
complement(3438. .5045)
/gene="fadd12"
/note="Rv1427c, (MTCY493.27), len: 535. fadd12, Possible
long chain long-chain-fatty-acid-CoA ligase similar to eg.
PC60_YEAST P38137 peroxisomal-coenzyme a synthetase (543
aa), fastascores; opt: 507 z-score: 579.4 E(): 2.9e-25,
30.4%identity in 365 aa overlap Contains PS00455 Putative
AMP-bindingdomain signature. Also similar to many M.
tuberculosisproteinseg. MTCY06A4.14 (44.8% identity in 525
aa overlap)"
/codon_start=1
/transl_table=11
/product="fadd12"
/protein_id="CAB09257.1"
/db_xref="GI:2131069"
/db_xref="SPTREMBL:006831"
/translation="MRIRQAFGLIATMRRAGLIAPLRPDRYLRIVAAAMREGMFTAG
FAGAARCPDRPGLIDELGTLTWKQDERGNALAAALQALPACPPRVVGINCRNHRGF
VDALIVNRIGAHITLLNTSFAGPALAEVVTREGVDTVVYDEFSATVDRALEKPOA
TRIVAWTDEDHDLTVEKLVAAAHAGRPETHGSHKVILLTSGTTGTPKGARHSGGIG
TLKAILDRTPWRAEEVTIVAPMFHAWGFSOLVLASSLACTIVTRRRFDPPEATLDLID
RHHATGLVVVPMFDRIMDLPAEIRNRYDGRSLRFAAASGSRMRPDVIAFMDFGDV
IYNNVNAIEAGMIATATPADIRAPDTAGPAEGTEIRIDQFTEPTGEVGTIYVR
NDSQFDGYTSGAAKDFHAGFMSSGDVGLDENGRLEFVVGGRDEMIIVSGENIYPIEVE
KTLATHPDYAEAAVIGVDDQYGGORLAAEVLKPGVSATPELTKQHVNDNLANKVPR
DIAVLDELPRGITGKILRTLEQSRVGS"
complement(4422. .4457)
/gene="fadd12"
/note="PS00455 Putative AMP-binding domain signature"
complement(5048. .5050)
/note="possible RBS upstream of Rv1427c"
complement(5049. .5876)
/gene="Rv1428c"
complement(5049. .5876)
/gene="Rv1428c"
/note="Rv1428c, (MTCY493.26), len: 275. Unknown, some
similarity to YV29_MYCTU Q1167 hypothetical 39.3 kd
protein cy20g9.29 (358 aa), fasta scores; opt: 355
z-score: 413.6 E(): 5e-16, 32.6% identity in 273 aa
overlap"
/codon_start=1
/transl_table=11
/product="hypothetical protein Rv1428c"
/protein_id="CAB09256.1"
/db_xref="GI:2131068"
/db_xref="SPTREMBL:006830"
/translation="MSETDSPGNDDAGIGDIGKFDPGLTQRLISVLRPVLKTYHRSQ
VHGLDSFPPGALVVAHNSGGMFPMDVPFVSVDYDKFGYDRPVYTLSDILFMGLTG
DLFRRTGIRATRENAAKALRSGGVVVPFGGDYDAYRPFAENVIDFNGRKGYVSTA
VEAGVPIVPAVSIGQESOLYLRGTWLAIRLGLRLRSLDILPISFGPFGFSAIP
PNLPIPAKIVMOVLDPINLTQGEDPDVDVADEHVRSVMOALNDLAAKRRFPILG"
complement(5882. .5885)
/note="possible RBS upstream of Rv1428c"
5982. .5985
/note="possible RBS upstream of Rv1429"
5995. .7263
/gene="Rv1429"
5995. .7263

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: January 31, 2003, 02:18:04 ; Search time 238.754 Seconds
(without alignments)
4584.089 Million cell updates/sec

Title: US-09-825-414-7
Perfect score: 2412
Sequence: 1 MHINRRVQQPPVTATDSFRT.....IEEGGTAASPSSEIPFPQRS 486

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ .p2n.model -DEV=xip
-Q=/cgn2_1/USPTO_spool/US09825414/runat_28012003_161741_26075/app_query.fasta_1.1294
-DB=N.Geneseq_101002 -QFMT=fastap -SUFFIX=p2n.rng -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blsum62 -TRANS=human40.cdl
-LIST=45 -DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09825414_@CGN_1_1_103_@runat_28012003_161741_26075 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV.TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N_Geneseq_101002:*

1: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2412	100.0	1461	22 AAD20408	P. syringae pv. tom
2	2412	100.0	1461	22 AAF55683	Nucleotide sequenc
3	2412	100.0	30365	22 AAD20405	P. syringae pv. to
4	1869.5	77.5	1464	22 AAD20438	Pseudomonas syring
5	168	7.0	4403765	22 AAI99683	Mycobacterium tube
6	164	6.8	4411529	22 AAI99682	Mycobacterium tube
7	148.5	6.2	4645	24 AAD31881	Lactobacillus rham
8	144.5	6.0	6744	13 AAQ29471	Extracellular fact
9	143.5	5.9	11883	21 AAA10263	Bordetella pertuss
10	143	5.9	1801	20 AAZ20203	Mycobacterium tube
11	142.5	5.9	7545	22 AAK51943	Human polynucleoti
12	142.5	5.9	7615	22 AAI58449	Human polynucleoti
13	142.5	5.9	7657	22 AAI58468	Human polynucleoti
14	141	5.8	2466	23 ABL27385	Drosophila melanog
15	141	5.8	53789	19 AAV21187	Amycolatopsis medi
16	139	5.8	1821	23 AAS81976	DNA encoding novel
17	138.5	5.7	4411529	22 AAI99682	Mycobacterium tube
18	137.5	5.7	2700	22 AAD02810	HfICU8 cDNA clone
19	137.5	5.7	12036	11 AAQ04668	FHA structural gen
20	137.5	5.7	32768	19 AAV52204	Streptococcus pneu
21	135.5	5.6	2550	19 AAV52497	Streptococcus pneu
22	135	5.6	2172	21 AAZ54239	Neisseria meningit
23	135	5.6	29069	21 AAA81497	N. meningitidis pa
24	135	5.6	349980	21 AAF21609	Neisseria meningit
25	134.5	5.6	4632	23 ABL25113	Drosophila melanog
26	134.5	5.6	9365	23 ABL25112	Drosophila melanog
27	133.5	5.5	2287	24 AAD28342	Mycobacterium spec
28	133.5	5.5	5036	22 AAD13450	Mycobacterium tube
29	132.5	5.5	1491	22 AAS59820	Human novel cytoxi
30	132.5	5.5	1924	24 ABK72305	Lymphona associate
31	132.5	5.5	1924	24 ABK72323	DNA encoding lymph
32	132.5	5.5	65140	22 AAD17184	Streptomyces nous
33	132.5	5.5	125401	22 AAD17186	Streptomyces nous
34	132.5	5.5	4403765	22 AAI99683	Mycobacterium tube
35	132	5.5	47981	22 AAF30757	Micromonospora meg
36	131.5	5.5	6258	23 AAS52002	Staphylococcus aur
37	131.5	5.5	17388	23 AAS54875	Staphylococcus aur
38	131.5	5.5	77536	21 ABA14651	Nucleotide sequenc
39	131	5.4	1800	24 ABA14137	DNA encoding antly
40	131	5.4	2916	23 ABL12469	Drosophila melanog
41	130.5	5.4	44377	18 AAT78508	Platenolide syntha
42	130.5	5.4	44377	18 AAT80414	Platenolide syntha
43	129.5	5.4	3606	23 ABL09625	Drosophila melanog
44	129.5	5.4	7035	23 AAS54978	Staphylococcus aur
45	129.5	5.4	9345	23 ABL09624	Drosophila melanog

ALIGNMENTS

RESULT 1	
AAD20408	
ID AAD20408	standard; DNA; 1461 BP.
XX AAD20408;	
XX	
DT 03-JAN-2002	(first entry)
DE P. syringae pv. tomato (Pto) DC3000	CEL open reading frame 5 (ORF5) DNA.
XX	
KW Conserved Effector Loc1; CEL; cytosolic; antibacterial; gene therapy;	
KW Exchangeable Effector Loc1; EEL; disease resistance; transgenic plant;	
KW eukaryotic cell death; cancer; ds.	
XX	
OS Pseudomonas syringae.	
XX	
PH Key	Location/Qualifiers
FT CDS	1..1461

FT /*tag= a
FT /product= "Pto DC3000 CEL ORF5 protein"
XX
PN W0200175066-A2.
XX 11-OCT-2001.
PD
PF 03-APR-2001; 2001W0-US10698.
XX
PR 03-APR-2000; 2000US-194160P.
PR 11-AUG-2000; 2000US-224604P.
PR 17-NOV-2000; 2000US-249548P.
XX
PA (CORR) CORNELL RES FOUND INC.
PA (UYNE-) UNIV NEVADA & COMMUNITY COLLEGE SYSTEM.
PA (UYNE-) UNIV NEBRASKA.
XX
PI Collier A, Alfano JR, Charkowski AO;
XX
DR WPI; 2001-639361/73.
DR P-PSDB; AAE12573.
XX
PT New nucleic acid molecules encoding proteins or polypeptides of
PT Pseudomonas Conserved Effector Loci and Exchangeable Effector Loci
PT genomic sequences, for imparting disease resistance to plants -
XX
PS Claim 1; Page 21; 217pp; English.
XX
CC The invention relates to an isolated nucleic acid molecule comprising a
CC nucleotide sequence encoding proteins or polypeptides of Pseudomonas
CC Conserved Effector Loci (CEL) and Exchangeable Effector Loci (EEL)
CC genomic sequences. CEL and EEL DNA are useful for imparting disease
CC resistance to a plant, by transforming a plant cell with the nucleic acid
CC and regenerating a transgenic plant from the transformed plant cell,
CC where the transgenic plant expresses a heterologous DNA molecule under
CC conditions effective to impart disease resistance, or by treating a plant
CC with an isolated protein or polypeptide, by applying the protein or
CC polypeptide in an isolated form or by applying a non-pathogenic bacteria
CC which secretes the protein or polypeptide, under conditions effective to
CC impart disease resistance to the treated plant. CEL and EEL proteins
CC are useful for causing eukaryotic cell death, by introducing a cytotoxic
CC pseudomonas protein into a eukaryotic cell under conditions effective to
CC cause cell death. CEL and EEL proteins are also useful for treating a
CC cancerous condition, by introducing a cytotoxic Pseudomonas protein into
CC cancer cells of a patient under conditions effective to cause death of
CC cancer cells, and thus treating the cancerous condition. The method
CC further involves administering a targeted DNA delivery system
CC comprising a DNA molecule encoding the cytotoxic Pseudomonas protein,
CC to the patient, where the targeted DNA delivery system delivers the
CC DNA molecule into cancer cells and the cytotoxic Pseudomonas protein
CC is expressed in the cancer cells. The present sequence is
CC Pseudomonas syringae pv. tomato (Pto) DC3000 CEL ORF5 DNA.
XX
SQ Sequence 1461 BP; 311 A; 440 C; 451 G; 259 T; 0 other;

Alignment Scores:
Pred. No.: 3.8e-169 Length: 1461
Score: 2412.00 Matches: 486
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 22 Gaps: 0

US-09-825-414-7 (1-486) x AAD20408 (1-1461)
QY 1 MethisIleAsnArgArgValGlnGlnProValThrAlaThrAspSerPheArgThr 20
Db 1 ATGCACATCAACCGACGCGTCCACAAACCGCCTGTGACTGCGACGGATAGCTTTCCGACA 60
QY 21 AlaserAspAlaserLeuAlaserSerSerValArgSerValSerSerAspGlnGlnArg 40
Db 61 GCGTCCGACGCGCTCTTGCTCCAGCTCTGTGCGATCTGTCAAGCTCGATCAACCAACGC 120

QY 41 GluIleAsnAlaIleAlaAspTyrLeuThrAspHisValPheAlaAlaHisLysLeuPro 60
Db 121 GAGATAAATGCGATTGCGGATTAACCTGACAGATCATGTGTTCGCTGGCATAAACTGCCG 180
QY 61 ProAlaaspSerAlaaspGlyGlnAlaAlaValaspValHisAsnAlaGlnIleThrAla 80
Db 181 CCGGCCGATTCGGCTGATGGCCCAAGCTGCAGTTGACGTACACAATGCGCAGATCACTGGC 240
QY 81 LeuIleGluThrArgAlaSerArgLeuHisPheGlnGlyGluThrProAlaThrIleAla 100
Db 241 CTGATCGAGACGCGCCGACCGCGCTGCACCTTCGAAGGGGAAACCCGGCAACCATCGCC 300
QY 101 AspThrPheAlaLysAlaGluLysLeuaspArgLeuAlaThrThrThrSerGlyAlaLeu 120
Db 301 GACACCTTCGCCAAGCGGGAAGCTCGACCGATTGGCGACGACTACATCAAGCGCGTTG 360
QY 121 ArgAlaThrProPheAlaMetAlaSerLeuLeuGlnTyrMetGlnProAlaIleAsnLys 140
Db 361 CGGGCGACGCCCTTTGCCATGGCCTCGTTCAGTACATGCAAGCTGCGATCAACAAG 420
QY 141 GlyAspTyrPleuProAlaProLeuLysProLeuThrProLeuIleSerGlyAlaLeuSer 160
Db 421 GCGGATTTGGCTGGCGGCTCCGCTCAAAACCGGTGAACCCCGCTCATTTCCGGAGCGCTGCG 480
QY 161 GlyAlaMetaspGlnValGlyThrLysMetMetaspArgAlaThrGlyAspLeuHisTyr 180
Db 481 GCGGCCATGGACAGGTGGGCACCAAGATGATGGACCGCGCAGCGGTGATCTGCATTAC 540
QY 181 LeuSerAlaSerProaspArgLeuHisaspAlaMetAlaAlaSerValLysArgHisSer 200
Db 541 CTGAGCGCTTCGCGGACAGGCTCCACGATGCGCGCTTCGGTGAAGCGCCACTCG 600
QY 201 ProSerLeuAlaArgGlnValLeuaspThrGlyValAlaValGlnThrThrSerAlaArg 220
Db 601 CCAAGCCTTGCTGTCAGACAGTTCTGACACAGCGGGTTCGGTTCAGACGTACTCGCGCGC 660
QY 221 AsnAlaValArgThrValLeuAlaProAlaLeuAlaSerArgProAlaValGlnGlyAla 240
Db 661 AACGCCGTACGTACCGTATTGGCTCCGGCACTGGCGCTCCAGACCCGCGGTGCAGGGTCT 720
QY 241 ValaspLeuGlyValSerMetAlaGlyGlyLeuAlaAlaAsnAlaGlyPheGlyAsnArg 260
Db 721 GTGGACCTTGGTGTATCGATGGCGGGGTGCTGCTGCCAACGCAAGCTTTGGCAACCGC 780
QY 261 LeuLeuSerValGlnSerArgaspHisGlnArgGlyGlyAlaLeuValLeuGlyLeuLys 280
Db 781 CTGCTCAGTGTGAGTTCGGGTGATCACCAAGCTGGCGGTGCATTAGTCTCGGTTTGAAG 840
QY 281 AspLysGluProLysAlaGlnLeuSerGlnGluAsnAspTyrPleuGluAlaTyrLysAla 300
Db 841 GATAAAGAGCCCAAGGCTCAACTGAGCGAAGAAACGACTGGCTCGAGCTTATAAAGCA 900
QY 301 IleLysSerAlaSerTyrSerGlyAlaAlaLeuAsnAlaGlyLysArgMetAlaGlyLeu 320
Db 901 ATCAATCGCGCCAGCTACTCGGGTGGCGGCTCAACGCTGGCAAGCGGATGGCCGCTCG 960
QY 321 ProLeuaspMetAlaThrAspAlaMetGlyAlaValArgSerLeuValSerAlaSerSer 340
Db 961 CCACTGGATATGCGGACCGACCGCAATGGGTGGGTAAAGAAGCCTGGTGTCAACGCTCAGC 1020
QY 341 LeuThrGlnAsnGlyLeuAlaLeuAlaGlyGlyPheAlaGlyValGlyLysLeuGlnGlu 360
Db 1021 CTGACCCAAAGCGTCTGGCCCTGGCGGGTGGCTTTGCAGGGGTAGCGAAGTTGCAGAGAG 1080
QY 361 MetAlaThrLysAsnIleThrAspProAlaThrLysAlaAlaValSerGlnLeuThrAsn 380
Db 1081 ATGGCGACGAATAATATCACCGACCGCGGCAAGCGCGGTCAAGTCAAGTTCAGCAAC 1140
QY 381 LeuAlaGlySerAlaAlaValPheAlaGlyTyrThrThrAlaAlaLeuThrThrAspPro 400
Db 1141 CTGGCAGGTTCCGACGCGCTTTTCGACAGGCTGGACCAACGCGCGCTGACAAACGATCCC 1200
QY 401 AlaValLysLysAlaGluSerPheIleGlnAspThrValLysSerThrAlaSerSerThr 420

|||||
Db 1201 GCGGTGAAAAAGCCGAGTGGTCATACAGACACGGTGAATCGACTGCATCCAGTACC 1260
QY 421 ThrGlyTyrValAlaAspGlnThrValLysLeuAlaLysThrValLysAspMetGlyGly 440
Db 1261 ACAGGCTACGTAGCCGACACCGCTCAAACTGGCGAAGACCGCTCAAGACATGGGCGG 1320
QY 441 GluAlaIleThrHisThrGlyAlaSerLeuArgAsnThrValAsnAsnLeuArgGlnArg 460
Db 1321 GAGCGGATCACCATACCGCGCAGCTTGCGCAATACGGTCAATACTGGCTCAACGC 1380
QY 461 ProAlaArgGluAlaAspIleGluGlyGlyThrAlaAlaSerProSerGluIlePro 480
Db 1381 CCGGCTCGTGAAGCTGATATAGAGAAGGGGGCAGCGCGGCTTCTCCAAGTGAATACCG 1440
QY 481 PheArgProMetArgSer 486
Db 1441 TTTCGGCCTATGCGGTCG 1458
RESULT 2
AAF55683
ID AAF55683 standard; DNA; 1461 BP.
XX AAF55683;
AC
XX 11-JUN-2001 (first entry)
DT
XX Nucleotide sequence of a HopPtoA protein of Pseudomonas syringae.
DE
XX HIV; tat protein; effector protein; transduction domain; HopPtoA protein;
KW ss.
KM Pseudomonas syringae.
XX
OS
XX Key Location/Qualifiers
FH CDS 1..1461
FT /*tag= a
FT /product= "HopPto"
FT
XX WO200119393-A1.
PN
XX 22-MAR-2001.
PD
XX 13-SEP-2000; 2000WO-US24977.
PF
XX 13-SEP-1999; 99US-0153507.
PR
XX (CORR) CORNELL RES FOUND INC.
PA
XX ColImer A, Beer SV;
PI
XX WPI: 2001-257850/26.
DR P-PSDB; AAB67677.
DR
XX
XX Delivering effector proteins into target cell for use in protein
PT therapy, involves introducing effector protein fused to protein
PT transduction domain of human immunodeficiency virus Tat protein, into
PT target cells -
PT
XX Example 2; Page 22-23; 43pp; English.
PS
XX The present sequence encodes a HopPtoA protein of Pseudomonas syringae
CC pv. tomato DC300 CEL.. This is an effector protein, which is used in
CC the method of the invention. The specification describes a method for
CC delivering effector proteins into a target cell. The method comprises
CC introducing an effector protein fused to a protein transduction domain
CC of a human immunodeficiency virus (HIV) tat protein into the target
CC cell. The method is used for delivering a heterologous effector protein
CC such as an effector protein produced by bacterial plant pathogen, animal
CC pathogen or a rhizosphere bacteria, or a protein secreted and/or
CC delivered into eukaryotic cells by a Type III secretion system or a
CC hypersensitive response elicitor, an avirulence protein, a
CC hypersensitive response and pathogenicity-dependent outer protein, a

CC virulence protein or a pathogenicity protein, into an eukaryotic cell.
CC The method is useful for delivering effector proteins for use in
CC pharmaceutical, insecticidal, fungicide, herbicide and other
CC applications.
CC
XX

SQ Sequence 1461 BP; 311 A; 440 C; 451 G; 259 T; 0 other;

Alignment Scores:
Pred. No.: 3.8e-169 Length: 1461
Score: 2412.00 Matches: 486
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 22 Gaps: 0

US-09-825-414-7 (1-486) x AAF55683 (1-1461)

QY 1 MethIleAsnArgArgValGlnGlnProProValThrAlaThrAspSerPheArgThr 20
Db 1 ATGCACATCAACCGACGCGCTCCACAACCGCCTGTGACTGCGACGGATAGCTTCCGACA 60
QY 21 AlaSerAspAlaSerLeuAlaSerSerValArgSerValSerSerAspGlnGlnArg 40
Db 61 GCGTCGACGCGCTCTTGGCTCCAGCTTGTGCGATGTGTCAAGCTCCGATCAGCAACGC 120
QY 41 GluIleAsnAlaIleAlaAspTyrLeuThrAspHisValPheAlaHisLysLeuPro 60
Db 121 GAGATAAATGCGATTGCCGATTAAGTGAAGATCATGTGTGCGCATAAACTGCCG 180
QY 61 ProAlaAspSerAlaAspGlyGlnAlaAlaValAlaSpValHisAsnAlaGlnIleThrAla 80
Db 181 CCGGCCGATTGCGCTGATGGCCAGCTGCAGTTGACGTACACATGCGCAGATCACTGCC 240
QY 81 LeuIleGluThrArgAlaSerArgLeuHisPheGluGlyGluThrProAlaThrIleAla 100
Db 241 CTGATCGAGACGCGCGCCAGCCGCTGCACCTTGAAGGGGAAACCCGGCAACCATCGCC 300
QY 101 AspThrPheAlaLysAlaGluLysLeuAspArgLeuAlaThrThrSerGlyAlaLeu 120
Db 301 GACACCTTCGCCAAGCGGAAAGCTCGACCGATTGGCGACACTACATCAGGCGCGTTG 360
QY 121 ArgAlaThrProPheAlaMetAlaSerLeuGlnTyrMetGlnProAlaIleAsnLys 140
Db 361 CCGGCGACGCCCTTGGCATGGCTCGTGTGCTTCACTACATGACGCTGCGATCAACAG 420
QY 141 GlyAspTrpLeuProAlaProLeuLysProLeuThrProLeuLysSerGlyAlaLeuSer 160
Db 421 GCGGATTGGCTGCCGCTCCGCTCAAAACCGCTGACCCCGCTCATTTCCGAGCGCTGCG 480
QY 161 GlyAlaMetAspGlnValGlyThrLysMetMetAspArgAlaThrGlyAspLeuHisTyr 180
Db 481 GCGGCCATGGACCAAGGTGGGCACCAAGATGATGACCGCGGCGGATGATCTGCATTAC 540
QY 181 LeuSerAlaSerProAspArgLeuHisAspAlaMetAlaAlaSerValLysArgHisSer 200
Db 541 CTGAGCGCTCGCCGCGACAGGCTCCACGATGCGATGGCCGCTTCGGTGAAGCGCACCTCG 600
QY 201 ProSerLeuAlaArgGlnValLeuAspThrGlyValAlaValGlnThrTyrSerAlaArg 220
Db 601 CCAAGCCTTGCTCGACAGGTTCTGACACAGGGGGTTGCGGTTCAAGACTCGGCGCGC 660
QY 221 AsnAlaValArgThrValLeuAlaProAlaLeuAlaSerArgProAlaValGlnGlyAla 240
Db 661 AACGCGTACGTACCGTATTGGCTCCGCACTGGCGTCCAGACCGCGGTGACAGGGTGCT 720
QY 241 ValAspLeuGlyValSerMetAlaGlyGlyLeuAlaAlaAsnAlaGlyPheGlyAsnArg 260
Db 721 GTGACCTTGGTGTATGATGAGCGGCTGCTGCTGCCAACGACGCTTGGCAACCGC 780
QY 261 LeuLeuSerValGlnSerArgAspHisGlnArgGlyGlyAlaLeuValLeuGlyLeuLys 280
Db 781 CTGCTCAGTGTGCAGTCCGCGTATACCAACGCGTGCGGCTGCATTAGTGCTCGGTTGAAG 840

QY 281 AsplysGluProLysAlaGlnLeuSerGluGluAsnAspTrpLeuGluAlaTyrLysAla 300
PI |
XX 841 GATAAAGAGCCCAAGGCTCACTGAGCGAAGAAAAACGACTGGCTCGAGGCTTATAAAGCA 900
QY 301 IleYsSerAlaSerTyrSerGlyAlaAlaLeuAsnAlaGlyLysArgMetAlaGlyLeu 320
DB |
DB 901 ATCAAATCGGCCAGCTACTCGGGTGCGCGCTCAACGCTGGCAAGCGGATGGCCGCTTG 960
QY 321 ProLeuAspMetAlaThrAspAlaMetGlyAlaValArgSerLeuValSerAlaSerSer 340
DB |
DB 961 CCACTGATATGGCGACCGACGCAATGGTGGGTGAAGACCTGTGTCAAGCTCCAGC 1020
QY 341 LeuThrGlnAsnGlyLeuAlaLeuAlaGlyGlyPheAlaGlyValGlyLysLeuGlnGlu 360
DB |
DB 1021 CTGACCCCAAAACGGTCTGGCCCTGGCGGGTGGCTTGAGGGGTAGGCAAGTTGACAGAG 1080
QY 361 MetaLathrLysAsnIleThrAspProAlaThrLysAlaAlaValSerGlnLeuThrAsn 380
DB |
DB 1081 ATGGCGACGAAAAATATCACCGACCCGCGACCAAGCGCGGTCACTCAGTTGACCAAC 1140
QY 381 LeuAlaGlySerAlaAlaValPheAlaGlyTyrThrThrAlaAlaLeuThrThrAspPro 400
DB |
DB 1141 CTGGCAGGTTCCGACGCCGTTTCCGACGGCTGACCAACGCCGCTGACACACCGATCCC 1200
QY 401 AlaValLysLysAlaGlySerPheIleGlnAspThrValLysSerThrAlaSerSerThr 420
DB |
DB 1201 GCGGTGAAAAAAGCCGAGTGTTCATACAGGACACGGTGAAATCGACTGCATCCAGTACC 1260
QY 421 ThrGlyTyrValAlaAspGlnThrValLysLeuAlaLysThrValLysAspMetGlyGly 440
DB |
DB 1261 ACAGGCTACGTAGCCGACGACCGCTCAAACTGGCGAAGACCGTCAAGACATGGCGGG 1320
QY 441 GluAlaIleThrHisThrGlyAlaSerLeuArgAsnThrValAsnAsnLeuArgGlnArg 460
DB |
DB 1321 GAGGCGATCACCCATACCGCGCCAGCTTGCGCAATACGGTCAATACCTGCGTCAACGC 1380
QY 461 ProAlaArgGluAlaAspIleGluGluGlyGlyThrAlaAlaSerProSerGluIlePro 480
DB |
DB 1381 CCGGCTCGTGAAGCTGATATAGAAGAGGGGGCAGCGGCTTCTCCAAGTGAATACCG 1440
QY 481 PheArgPrometArgSer 486
DB |
DB 1441 TTTCCGGCCTATGCGGTCG 1458
RESULT 3
AAD20405/C
ID AAD20405 standard; DNA; 30365 BP.
XX
AC AAD20405;
XX
DT 03-JAN-2002 (first entry)
XX
DE P. syringae pv. tomato (Pto) DC3000 Conserved Effector Loci (CEL) DNA.
XX
KW Conserved Effector Loci; CEL; cytostatic; antibacterial; gene therapy;
KW Exchangeable Effector Loci; EEL; disease resistance; transgenic plant;
XX eukaryotic cell death; cancer; ds.
OS Pseudomonas syringae.
XX
PN WO200175066-A2.
XX
PD 11-OCT-2001.
XX
PF 03-APR-2001; 2001WO-US10698.
XX
PR 03-APR-2000; 2000US-194160P.
PR 11-AUG-2000; 2000US-224604P.
PR 17-NOV-2000; 2000US-249548P.
XX
PA (CORR) CORNELL RES FOUND INC.
PA (UYNE-) UNIV NEVADA & COMMUNITY COLLEGE SYSTEM.
PA (UYNE-) UNIV NEBRASKA.

XX Colimer A, Alfano JR, Charkowski AO;
PI
XX WPI: 2001-639361/73.
DR
XX
XX New nucleic acid molecules encoding proteins or polypeptides of
PT Pseudomonas Conserved Effector Loci and Exchangeable Effector Loci
PT genomic sequences, for imparting disease resistance to plants
XX
XX
PS Disclosure; Page 9-17; 217pp; English.
XX
XX The invention relates to an isolated nucleic acid molecule comprising a
CC nucleotide sequence encoding proteins or polypeptides of Pseudomonas
CC Conserved Effector Loci (CEL) and Exchangeable Effector Loci (EEL)
CC genomic sequences. CEL and EEL DNA are useful for imparting disease
CC resistance to a plant, by transforming a plant cell with the nucleic acid
CC and regenerating a transgenic plant from the transformed plant cell,
CC where the transgenic plant expresses a heterologous DNA molecule under
CC conditions effective to impart disease resistance, or by treating a plant
CC with an isolated protein or polypeptide, by applying the protein or
CC polypeptide in an isolated form or by applying a non-pathogenic bacteria
CC which secretes the protein or polypeptide, under conditions effective to
CC impart disease resistance to the treated plant. CEL and EEL proteins
CC are useful for causing eukaryotic cell death, by introducing a cytotoxic
CC Pseudomonas protein into a eukaryotic cell under conditions effective to
CC cause cell death. CEL and EEL proteins are also useful for treating a
CC cancerous condition, by introducing a cytotoxic Pseudomonas protein into
CC cancer cells of a patient under conditions effective to cause death of
CC cancer cells, and thus treating the cancerous condition. The method
CC further involves administering a targeted DNA delivery system
CC comprising a DNA molecule encoding the cytotoxic Pseudomonas protein,
CC to the patient, where the targeted DNA delivery system delivers the
CC DNA molecule into cancer cells and the cytotoxic Pseudomonas protein
CC is expressed in the cancer cells. The present sequence is
CC Pseudomonas syringae pv. tomato (Pto) DC3000 CEL DNA.
XX
SQ Sequence 30365 BP; 6828 A; 8938 C; 8516 G; 6080 T; 3 other;
Alignment Scores:
Pred. No.: 1.32e-167 Length: 30365
Score: 2412.00 Matches: 486
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 22 Gaps: 0
US-09-825-414-7 (1-486) x AAD20405 (1-30365)
QY 1 MethisIleAsnArgArgValGlnGlnProProValThrAlaThrAspSerPheArgThr 20
DB |
DB 23984 ATGCACATCAACCGACGCGTCCAAACACCGCCTGTACTGCGACGATAGCTTCGGACA 23925
QY 21 AlaserAspAlaSerIleuAlaSerSerSerValArgSerValSerSerAspGlnArg 40
DB |
DB 23924 GCGTCCGACGCGCTCTGCTCCACAGCTCTGCGATCTGTACAGTCCGATCAGCAACGC 23865
QY 41 GluIleAsnAlaIleAlaAspTyrLeuThrAspHisValPheAlaAlaHisLysLeuPro 60
DB |
DB 23864 GAGATAAATGCGATTGCCGATTACCTGACAGATCATGTGTTCGTCGCCATTAACGCGC 23805
QY 61 ProAlaAspSerAlaAspGlyGlnAlaAlaValAspValHisAsnAlaGlnIleThrAla 80
DB |
DB 23804 CCGGCCGATTGGCTGATGGCCAAAGCTGCAGTTGACGTACACAATGGCGCAGATCAGTGGC 23745
QY 81 LeuIleGluThrArgAlaSerArgLeuHisPheGluGlyGluThrProAlaThrIleAla 100
DB |
DB 23744 CTGATCGAGACGCGCGCCAGCCCTGCACCTTGAAGGGAACCCCGCAACCATCGCC 23685
QY 101 AspThrPheAlaLysAlaGluLysLeuAspArgLeuAlaThrThrThrSerGlyAlaLeu 120
DB |
DB 23684 GACACCTTCGCCCAAGCGGGAAGAGCTGCAGCCGATGGCGACGACTACATCAGCGCGTTG 23625
QY 121 ArgAlaThrProPheAlaMetAlaSerLeuLeuGlnTyrMetGlnProAlaIleAsnLys 140

Db 23624 CGGGCGACGCCCTTGGCCATGGCCTCGTTGCTTCAGTACATGCAGCCCTGGCATCAACAAG 23565
QY 141 GLYAspTrPLeuProAlaPrLeuLysProLeuThrProLeuIleSerGlyAlaLeuSer 160
Db 23564 GCGGATTGGCTGCCGCTCCGCTCAAAACCGCTGACCCCGCTCATTTCCGGAGCGCTGTCG 23505
QY 161 GlyAlaMetAspGlnValGlyThrLysMetMetAspArgAlaThrGlyAspLeuHisTyr 180
Db 23504 GCGGCCATGGACGAGTGGGACCAAGATGATGACCGCGCAGCGGGTGATCTGCATTAC 23445
QY 181 LeuSerAlaSerProAspArgLeuHisAspAlaMetAlaAlaSerValLysArgHisSer 200
Db 23444 CTGAGCGCCTCGCCGACAGCCTCCACGATGCGATGGCCGCTTCGGTGAAGCCCACTCG 23385
QY 201 ProSerLeuAlaArgGlnValLeuAspThrGlyValAlaValGlnThrTyrSerAlaArg 220
Db 23384 CCAAGCCTTGCTCGACAGGTTCTGGACACGGGGGTTGCCGTTCAAGACGTACTCGCGCGC 23325
QY 221 AsnAlaValArgThrValLeuAlaProAlaLeuAlaSerArgProAlaValGlnGlyAla 240
Db 23324 AACGCCGTACGTACCGTATTGGCTCCGGCACTGGCGCTCCAGACCCCGCGTGACGGTGCT 23265
QY 241 ValAspLeuGlyValSerMetAlaGlyGlyLeuAlaAlaAsnAlaGlyPheGlyAsnArg 260
Db 23264 GTGGACCTTGGTATGATGATGGCGGGGTGCTGTGGCTGCCCAACGACAGCCTTGGCAACGCC 23205
QY 261 LeuLeuSerValGlnSerArgAspHisGlnArgGlyAlaLeuValLeuGlyLeuLys 280
Db 23204 CTGCTCAGTGTGCAGTCCGCTGATCACACGACGTGGCGGTGCATTAGTCTCGGTTGAAG 23145
QY 281 AspLysGluProLysAlaGlnLeuSerGlnGluAsnAspTrpLeuGluAlaTyrLysAla 300
Db 23144 GATTAAGAGCCCAAGGCTCACTGAGCGAAGAAAACGACTGGCTCGAGGCTTATAAGCA 23085
QY 301 IleLysSerAlaSerTyrSerGlyAlaAlaLeuAsnAlaGlyLysArgMetAlaGlyLeu 320
Db 23084 ATCAAAATCGCCACGCTACTCGGGGTGGCGCGCTCAACGCTGCCAAGCGGATGGCCGGTCTG 23025
QY 321 ProLeuAspMetAlaThrAspAlaMetGlyAlaValArgSerLeuValSerAlaSerSer 340
Db 23024 CCACTGGATATGGCGACCGACCAATGGGTGCCGTTAAGAACCTGTGTACAGCCTCCAGC 22965
QY 341 LeuThrGlnAsnGlyLeuAlaLeuAlaGlyGlyPheAlaGlyValGlyLysLeuGlnGlu 360
Db 22964 CTGACCCCAAAACGGTCTGGCCCTGGCGGGGTGGCTTGGCAGGGGTAGGCCAAGTTGCAGAG 22905
QY 361 MetAlaThrLysAsnIleThrAspProAlaThrLysAlaAlaValSerGlnLeuThrAsn 380
Db 22904 ATGGCGACGAAAAATATACCCGACCCGGCGACCAAGCGCGGTCAGTCAAGTTGACCAAC 22845
QY 381 LeuAlaGlySerAlaAlaValPheAlaGlyTyrTrpThrAlaAlaLeuThrThrAspPro 400
Db 22844 CTGGCAGGTTGCGCAGCCGTTTTCGACGCTGACACACGGCCGCTGACACACCGATCCC 22785
QY 401 AlaValLysLysAlaGluSerPheIleGlnAspThrValLysSerThrAlaSerSerThr 420
Db 22784 GCGGTGAAAAAAGCCGAGTCTTCATACAGACACCGGTGAATCGACTGCATCCAGTACC 22725
QY 421 ThrGlyTyrValAlaAspGlnThrValLysLeuAlaLysThrValLysAspMetGlyGly 440
Db 22724 ACAGGCTACGTAGCCGACCAAGACCGTCAAACTGGCGGAAGACCGTCAAAAGACATGGCGGG 22665
QY 441 GluAlaIleThrHisThrGlyAlaSerLeuArgAsnThrValAsnAsnLeuArgGlnArg 460
Db 22664 GAGGCGATCACCCATACCGGGCCAGCTTGGCAATACGTTCAATTAACCTGCGTCAACGC 22605
QY 461 ProAlaArgGluAlaAspIleGluGluGlyGlyThrAlaAlaSerProSerGluIlePro 480
Db 22604 CCGGCTCGTGAAGCTGATATGAAAGAGGGGGGCACGGCGCTTCTCCAAGTGAATATACCG 22545
QY 481 PheArgPrometArgSer 486
|||||

Db 22544 TTTCGGCCTATGCGGTGCG 22527
RESULT 4
AAD20438
ID AAD20438 standard; DNA; 1464 BP.
XX
AC AAD20438;
XX
DT 03-JAN-2002 (first entry)
XX
DE Pseudomonas syringae pv. tomato strain DC3000 HopPtoA2 DNA.
KW Conserved Effector Loci; CEL; cytostatic; antibacterial; gene therapy;
KW Exchangeable Effector Loci; EEL; disease resistance; transgenic plant;
KW eukaryotic cell death; cancer; ds.
XX
OS Pseudomonas syringae.
XX
FH
FT
FT CDS Location/Qualifiers
1..1464
/*tag= a
/product= "P. syringae pv. tomato HopPtoA2 protein"
WO200175066-A2.
XX
PD 11-OCT-2001.
XX
PE 03-APR-2001; 2001WO-US10698.
XX
PR 03-APR-2000; 2000US-194160P.
PR 11-AUG-2000; 2000US-224604P.
PR 17-NOV-2000; 2000US-249548P.
XX
PA (CORR) CORNELL RES FOUND INC.
PA (UYNE-) UNIV NEVADA & COMMUNITY COLLEGE SYSTEM.
XX
PI (UYNE-) UNIV NEBRASKA.
XX
PI COLLIER A, Alfano JR, Charkowski AO;
XX
DR WPI; 2001-639361/73.
XX
DR P-PSDB; AAE12603.
XX
PT New nucleic acid molecules encoding proteins or polypeptides of
PT Pseudomonas Conserved Effector Loci and Exchangeable Effector Loci
XX genomic sequences, for imparting disease resistance to plants -
XX
PS Claim 1; Page 64; 217pp; English.
XX
CC The invention relates to an isolated nucleic acid molecule comprising a
CC nucleotide sequence encoding proteins or polypeptides of Pseudomonas
CC Conserved Effector Loci (CEL) and Exchangeable Effector Loci (EEL)
CC genomic sequences. CEL and EEL DNA are useful for imparting disease
CC resistance to a plant, by transforming a plant cell with the nucleic acid
CC and regenerating a transgenic plant from the transformed plant cell,
CC where the transgenic plant expresses a heterologous DNA molecule under
CC conditions effective to impart disease resistance, or by treating a plant
CC with an isolated protein or polypeptide, by applying the protein or
CC polypeptide in an isolated form or by applying a non-pathogenic bacteria
CC which secretes the protein or polypeptide, under conditions effective to
CC impart disease resistance to the treated plant. CEL and EEL proteins
CC are useful for causing eukaryotic cell death, by introducing a cytotoxic
CC Pseudomonas protein into a eukaryotic cell under conditions effective to
CC cause cell death. CEL and EEL proteins are also useful for treating a
CC cancerous condition, by introducing a cytotoxic Pseudomonas protein into
CC cancer cells of a patient under conditions effective to cause death of
CC cancer cells, and thus treating the cancerous condition. The method
CC further involves administering a targeted DNA delivery system
CC comprising a DNA molecule encoding the cytotoxic Pseudomonas protein,
CC to the patient, where the targeted DNA delivery system delivers the
CC DNA molecule into cancer cells and the cytotoxic Pseudomonas protein
CC is expressed in the cancer cells. The present sequence is a DNA
CC encoding Pseudomonas syringae pv. syringae HopPysA homolog protein.
XX

Sequence 1464 BP; 325 A; 392 C; 442 G; 305 T; 0 other;

Alignment Scores:

Pred. No.:	4 27e-129	Length:	1464
Score:	1869.50	Matches:	381
Percent Similarity:	85.19%	Conservative:	33
Best Local Similarity:	78.40%	Mismatches:	71
Query Match:	77.51%	Indels:	1
DB:	22	Gaps:	1

US-09-825-414-7 (1-486) x AAD20438 (1-1464)

QY		1	MechHisIleasnArgValGlnGlnProProValThrAlaThrAspSerPheArgThr	20
Db		1	ATGCACATCAACCAATCCGCCCAACAACCGCTGGCGTTGCAATGAGAGTTCGGACA	60
QY		21	AlaserAspaLaserLeuAlaserSerSerValArgSerValSerSerAspGlnArg	40
Db		61	GCTTCCGACGCGTCCCTTGCTGCAgTCTGTGGCGGTCTGTCAAGCATACTACTGTCGCCG	120
QY		41	GluiIleasnaIlealeaAspTyrLeuThrAspHisValPheAlaAlaHisLysLeuPro	60
Db		121	GATCTACAAGCTATTACCgATTATCTGAACAATCACGTGTTCGCTGCCACAGCTTTTCG	180
QY		61	ProAlaaspSerAlaaspGlyGlnAlaAlaValaspValHisAsnAlaGlnIleThrAla	80
Db		181	GTAATAGGCTCACCGGATGAGCGGTGATGCCGCTCTTGACACACAAGCAGCATCGATCGCG	240
QY		81	IeuIleGluThrArgAlaSerArgLeuHisPheGluGlyGluThrProAlaThrIleAla	100
Db		241	TTCGTAGAGACACGCGCCAACCGCTGTACTCCGAAGGGGAGACCCCCGCAACCATCGCC	300
QY		101	AspThrPheAlaLysAlaGluLysLeuaspArgLeuAlaThrThrThrSerGlyAlaLeu	120
Db		301	GAAACATTCGCCAAGCGGAAGAAGTTCGACCCTTTGGCGACGACCGCATCAAGTGTCTTT	360
QY		121	ArgAlaThrProPheAlaMetAlaSerLeuLeuGlnTyrMetGlnProAlaIleasnLys	140
Db		361	GAGAACACGCCCATTTGCCCGCTCGCTGCTTCAGTACATGCAGCCTCGCATCAACAAG	420
QY		141	GlyAspTrpleuProAlaProLeuLysProLeuThrProLeuIleSerGlyAlaLeuSer	160
Db		421	GGCGATTGGCTAGCAACGCCGCTCAAGCCGCTGACCCCGCTCATTTCCGGACGCTGTGC	480
QY		161	GlyAlaMetaspGlnValGlyThrLysMetMetaspArgAlaThrGlyaspLeuHisTyr	180
Db		481	GGAGCCATGGACCAAGGTGGGCGACCAAAATGATGGATCGTGCAGAGGGGTGATTCGATTAC	540
QY		181	LeuSerAlaSerProaspArgLeuHisaspAlaMetAlaAlaSerValLysArgHisSer	200
Db		541	CTGAGCATTTCGCCCGACACAAGTTCATGATGCGATGGCCGTATCGGTGAAGGCCCACTCG	600
QY		201	ProSerLeuAlaArgGlnValLeuaspthrGlyValAlaValaGlnThrTyrSerAlaArg	220
Db		601	CCTGCGCTTGGTCGACAGAGTGTGTGACATGGGGGATTGCAAGTCAGACGTTCTCGGCGCTA	660
QY		221	AsnaIaValaArgThrValLeuAlaProAlaLeuAlaSerArgProAlaValaGlnGlyAla	240
Db		661	AATGTGGTGCCTACCGTATTGGCTCCAGCACACTAGCGTCCAGACCCGTCCGGTGCAGGGTGCT	720
QY		241	ValaspLeuGlyValSerMetalaGlyGlyLeuAlaAlaAsnAlaGlyPheGlyAsnArg	260
Db		721	GTTGATTTTGGCGTATCTACGGCGGGGTGGCTTGGTTGCGAATGCAAGCTTTGGCGACCGC	780
QY		261	LeuLeuSerValGlnSerArgaspHisGlnArgGlyGlyAlaLeuValLeuGlyLeuLys	280
Db		781	ATGCTCAGTGTGCAATCGCGCGATCAACTGCGTGGGGGGGCATTCGTACTTGGCATGAAA	840
QY		281	AsplysgLuProLysAlaGlnLeuSerGluGluAsnAspTrpleuGlnAlaTyrLysAla	300
Db		841	GATAAAGAGCCCAAGCGCGCGTGTGACTGAAGAACTGATTGGCTTGATGCTTTACAAAGCG	900
QY		301	IleLysSerAlaSerTyrSerGlyAlaAlaLeuAsnAlaGlyLysArgMetAlaGlyLeu	320

Db	901	ATCAAGTCGGCCAGCTACTACAGTGGCGGCGCTCAATGCGGGCAAGCGATGGCCGGCTG	960
QY	321	ProLeuAspMetAlaThrAspAlaMetGlyAlaValArgSerLeuValSerAlaSerSer	340
Db	961	CCACTGGACGCTCGCGACCGCGCTCAAGGCGGTGAGAGTCTGTGTCCGCCACCAAGC	1020
QY	341	LeuThrGlnAsnGlyLeuAlaLeuAlaGlyGlyPheAlaGlyValGlyLysLeuGlnGlu	360
Db	1021	CTGACCAAAAAATGGCCCTGCGCTAGCCGGTGTGTACGCCCGGGGTAAAGTTGCAGAAA	1080
QY	361	MetaIaThrLysAsnIleThrAspProAlaThrLysAlaAlaValSerGlnLeuThrAsn	380
Db	1081	ATGGCGACGAAAAATATACACTGATTCGGCGACCAAGGCTGGCGTTAGTCAGCTGAGCAAC	1140
QY	381	LeuAlaGlySerAlaAlaValPheAlaGlyTrpThrThrAlaAlaLeuThrThrAspPro	400
Db	1141	CTGGTGGGTTTCGGTAGCGGTTTCGCACAGGCTGGACCACCGCTGGACTGGCGACTGACCT	1200
QY	401	AlaValLysLysAlaGluSerPheIleGlnAspThrValLysSerThrAlaSerSerThr	420
Db	1201	GCGGTTAAGAAAGCCGAGTCGTTTATACAGGATTAAGGTGAATCGAACCGCATCTAGTACC	1260
QY	421	ThrGlyTyrValAlaAspGlnThrValLysLeuAlaLysThrValLysAspMetGlyGly	440
Db	1261	ACAAGCTATGTTGCCGACCAAGCCGCTCAAACTGGCGAAACAGTCAAGGACATGAGCGGG	1320
QY	441	GluAlaIleThrHisThrGlyAlaSerLeuArgAsnThrValAsnLeuArgGlnArg	460
Db	1321	GAGCGCATCTCCAGCACCGGTCCAGCTTACGCGAGTACTGCAATAAACCCTGCGTCATCGC	1380
QY	461	ProAlaArgGluAlaAspIleGluGluGlyGlyThrAlaAla---SerProSerGluIle	479
Db	1381	TCCGCTCCGGAAGCTGATATCGAAGAAGTGGGATTTTCGGCGTTTCTCGAAGTGAACA	1440
QY	480	PropheArgPrometArg 485	
Db	1441	CCGTTTCAGCTCAGCGCT 1458	
RESULT 5			
AA199683/C			
ID	AA199683	standard; DNA; 4403765 BP.	
XX	AC	AA199683;	
XX	DT	15-JAN-2002 (first entry)	
DE	Mycobacterium tuberculosis strain H37Rv genome SEQ ID NO 2.		
KW	Mycobacterium tuberculosis; strain H37Rv; strain CDC 1551; genome;		
KW	variation; epidemiology; patient treatment; epidemic monitoring; ds.		
XX	OS	Mycobacterium tuberculosis.	
XX	PN	US6294328-B1.	
XX	PD	25-SEP-2001.	
XX	PF	24-JUN-1998; 98US-0103840.	
XX	PR	24-JUN-1998; 98US-0103840.	
PA	(GENO-) INST GENOMIC RES.		
PI	Fleischmann RD, White OR, Fraser CM, Venter JC;		
DR	WPI; 2001-647261/74.		
XX	Evaluating strain variation of Mycobacterium tuberculosis, comprises		
PT	determining the nucleotide sequence of the strain at positions in the		
PT	genome corresponding to positions where M. tuberculosis strains CDC		
PT	1551 and H37Rv differ -		
XX			

CC H37Rv (AAI99682). The method is useful for evaluating strain variation of
CC M. tuberculosis and has valuable application in the fields of
CC tuberculosis genetics, epidemiology, patient treatment and epidemic
CC monitoring.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from USPTO
CC at seqdata.uspto.gov/sequence.html?docid=6294328B1.

XX Sequence 4411529 BP; 758565 A; 1449983 C; 1444602 G; 758379 T; 0 other;

Alignment Scores:

Pred. No.:	40.7	Length:	4411529
Score:	164.00	Matches:	95
Percent Similarity:	37.40%	Conservative:	43
Best Local Similarity:	25.75%	Mismatches:	133
Query Match:	6.80%	Indels:	98
DB:	22	Gaps:	14

US-09-825-414-7 (1-486) x AAI99682 (1-4411529)

QY 145 ProAlaProLeuLysProLeuThrProLeuIleSer--GlyAlaLeuSerGlyAlaMeta 164
Db 3943125 CCGGCGCCG-----CCGCTGCCCGCGCTTGAGTGAAGAGGCTGCCGCCGCCGCCG 3943072

QY 164 spGlnValGlyThrLysMetMetaspargAlaThrGlyAspLeuHisTyrLeuSerAlas 184
Db 3943071 GTGCCGCCGCGCACCGCTTGCGCGCGCTTGCCACGCGT----- 3943034

QY 184 erProAspArgLeuHisAspAlaMetaAlaAlaSerValLysArgHisSerProSerLeuA 204
Db 3943033 -----GCCGTAGCGCCGCCACCGCGCGCGCCAGCCCGGC---- 3942998

QY 204 laArgGlnValLeuAspThrGlyValAlaValGlnThrTyrSerAlaArgAsnAlaValA 224
Db 3942997 -----CGCCCCACCGGACACAC----- 3942980

QY 224 rgThrValLeuAlaProAlaLeuAlaSerArgProAlaValGlnGlyAlaValasPleuG 244
Db 3942979 -----GCTGCCGCTTGACCTCTTGCCCGCGCTTGCGCGCGGACTGCTGTG 3942931

QY 244 lyValSerMetaIaGlyGlyLeuAlaAlaAsnAlaGlyPheGlyAsnArg----- 260
Db 3942930 GGCCCGCT--GCCGCGCGCGCGCGCGCGCGACCGCGACTTCGCCGTCGCCGCTGCGC 3942874

QY 261 --LeuLeuSerValGlnSerArgAspHisGlnArgGlyAlaLeuValLeuGlyLeuL 280
Db 3942873 CCCTTGCCGCGCGCGCACCGCACCGATACCGTCGCCCGCGCGCGCACCACT----- 3942824

QY 280 ysAspLysGluProLysAlaGlnLeuSerGluGluAsnAspTrpLeuGluAlaTyrLysA 300
Db 3942824 ----- 3942824

QY 300 laIleLysSerAlaSerTyrSerGlyAlaAlaLeuAsnAlaGlyLysArgMetaIaGlyL 320
Db 3942823 -----GCCGCTCTGCGCGCGCGCTGCGCGCTGCGCGCGAGCGCGGTGCCGCC 3942772

QY 320 euProLeuAspMetaIaThrAspAlaMetGly---AlaValArgSerLeuValSerAlas 339
Db 3942771 CCACCGTTGCTGCAGCCGTGGCGGCGCGCGCGCGCGCGCGCTTGCGCGCGTCCGCCG 3942712

QY 339 erSerLeuThrGlnAsnGlyLeuAlaLeuAlaGly---GlyPheAlaGlyValGlyLysL 358
Db 3942711 GTACCGTTGGCGCGCGCACCGGTACCGCGCTGCTGCCCGCGCGCGCACCGCGCGGC---- 3942656

QY 358 euGlnGluMetaIaThrLysAsnIleThrAspProAlaThrLysAlaAlaValSerGlnL 378
Db 3942655 -----GTTTCGCCCTTTCGCCCAATCCACCGCTTTCGCCGCTATTTCGCTA 3942607

QY 378 euThrAsnLeuAlaGlySerAlaAlaValPheAlaGlyTrpThrThrAlaAlaLeuThrt 398
Db 3942606 CCACCGGCGTCTCCGCCACGCGCGCGATGCGCGCTTGCGCGCGTGCCTGCCGCGGTACT 3942547

QY 398 hrAspProAlaValLysLysAlaGluSerPheIleGlnAspThrValLysSerThr---- 416

Db 3942546 CCGGTCCCGCGCGCGAGCGCGTGCATGCGCGCTACCGCACGACCGCGGTTCACCGTTG 3942487

QY 417 -----AlaSerSerThrThrGlyTyrValA 425

Db 3942486 CCGCGCGCGCGCGGTTCGCCCATGCGCGCGCGCTGCCACAGCGCGGTGCCGAGATG 3942427

QY 425 laAspGlnThrValLysLeuAlaLysThrValLysAspMetGlyGlyAlaIle---- 443

Db 3942426 CCGACGGCACCGCGCGCATCTCCGCCGTGCCCATTTGCCACGATACCGTTGGCG 3942367

QY 444 -----ThrHisThrGlyAlaSerLeuArgAsnThrValAsnAsnLeuArgGlnArgProA 462

Db 3942366 CCGGAGTGGCACCGCGCGCGC--CGCGCGCGGTCTCCGCCCTTCCGCCGCTGCCG 3942310

QY 462 laArgGluAlaAspIleGluGly-----GlyThrAlaAlaSerP 476

Db 3942309 CCCGCGCGCGCGCTTGCGCGCGGTGCTGCTGTGACGCGCTGCTCCACCGTCGCGCCT 3942250

QY 476 roSerGluIleProPheArgPro 483

Db 3942249 TTGCCGCATTCGCCCGCGCGCC 3942227

RESULT 7

AAD31881

ID AAD31881 standard; DNA; 4645 BP.

XX AAD31881;

AC AAD31881;

XX 18-JUN-2002 (first entry)

DE Lactobacillus rhamnosus outer membrane protein rompa gene.

XX

KW Enzyme; flavour; aroma; texture; nutritional; dairy manufacture; therapy;
KW fermentation process; anti-infection; rotavirus infection; heart disease;
KW infantile diarrhoea; lactose digestion; anti-cancer; autoimmune disorder;
KW anti-mutagenesis; immune system modulation; allergy; Helicobacter pylori;
KW antihypertensive effect; urogenital infection; hepatic encephalopathy;
KW bowel syndrome; endocarditis; transgenic microbe; outer membrane protein;
KW rompa gene; ds.

XX

OS Lactobacillus rhamnosus HN001.

XX

XX

FX Key Location/Qualifiers

FT CDS 74..4465

FT /*tag= a

FT /product= "Outer membrane protein"

FT /note= "CDS does not include start codon"

FT /partial

XX

PN WO200212506-A1.

XX

PD 14-FEB-2002.

XX

PF 08-AUG-2001; 2001WO-NZ00160.

XX

PR 08-AUG-2000; 2000US-0634238.

XX

PR 28-NOV-2000; 2000US-0724623.

XX

PA (GENE-) GENESIS RES & DEV CORP LTD.

XX (VIAL-) VIALACTIA BIOSCIENCE NZ LTD.

PI Glenn M, Havukala J, Bloksberg LN, Lubbers MW, Dekker J;

PI Christensson AC, Holland R, O'toole PW, Reid JR, Coolbear T;

XX

DR WPT; 2002-241760/29.

DR P-PSDB; AAE20110.

XX

PT New polynucleotides and polypeptides from Lactobacillus rhamnosus,
PT useful in e.g. improving the flavor, aroma, texture and health-related
PT benefits of milk-derived products, or in increasing properties of
PT microbes -

XX

PS Claim 2; Fig 63; 257pp; English.

xx The present invention relates to a new isolated polynucleotide comprising a sequence present in Lactobacillus rhamnosus strain HN001 and encoding a polypeptide capable of modifying the flavour, aroma, texture, nutritional and health benefits of milk-derived products, and/or survivability of microbes in dairy manufacturing processes. The polynucleotides are useful for improving the properties of microbes used in the manufacture of milk-derived products such as cheeses, yogurt, fermented milk products, sour milks and buttermilk; in modifying the flavour, aroma, texture and health -related benefits of milk-derived products and in increasing the survival of microbes during industrial fermentation processes. The bacteria may be used to increase resistance to enteric pathogens and anti-infection activity, including treatment of rotavirus infection and infantile diarrhoea; aid in lactose digestion; as anti-cancer and anti-mutagenesis; liver cancer reduction; reduction of small bowel bacterial overgrowth; immune system modulation and treatment of autoimmune disorders and allergies; treatment of allergic responses to foods; reduction of blood lipids and prevention of heart disease; antihypertensive effect; prevention and treatment of urogenital infections, Helicobacter pylori, or hepatic encephalopathy; treatment of inflammatory bowel disorder and irritable bowel syndrome; modulation of endocarditis; and for improved protein and carbohydrate utilization and conversion. The transgenic microbial population can be administered to a mammal as an anti-carcinogenic agent. The present sequence is Lactobacillus rhamnosus outer membrane protein rompa gene.

xx Sequence 4645 BP; 1279 A; 1062 C; 1261 G; 1043 T; 0 other;

Alignment Scores:

Pred. No.: 0.186 Length: 4645
Score: 148.50 Matches: 126
Percent Similarity: 36.18% Conservative: 69
Best Local Similarity: 23.38% Mismatches: 233
Query Match: 6.16% Indels: 112
DB: 24 Gaps: 21

US-09-825-414-7 (1-486) x AAD31881 (1-4645)

QY 13 ThrAlaThrAspSerPheArgThrAla-----SerAspAlaSerLeuAlaSerSer--- 29
Db 2864 ACTGCACATCTGCGTATGCTAGTGTGCAAGTTCGAGTGCACGTGAAGCCGGTTCATAT 2923
QY 30 -----SerValArgSerValSerSerAspGlnGlnArgGluIleAsnAlaIleAlaAsp 47
Db 2924 GCTCATCAGGCGAGGCTCCAGCGCCAGTGCAGCTGTGCTAGTCCGGCAGTGCAGCCCA 2983
QY 48 TyrLeuThrAspHisValPheAlaAlaHisLysLeuProPro-----AlaAspSer 64
Db 2984 CATGCCAGCACCGCTGCGAGTGCAGTCCAGTATCCGAAGGATAGTGGATTCAGTCA 3043
QY 65 AlaAspGlnAlaAlaValAspValHisAsnAlaGlnIleThrAlaLeuIleGluThr 84
Db 3044 CTAGCCAGTCAAGGCTGCAAGCGGACGCAAGGCAAGCAGTAACGCGAGTGCCTCCCAACC 3103
QY 85 ArgAlaSerArgLeuHisPheGluGluThrProAlaThr---IleAlaAspThrPhe 103
Db 3104 AGCGCCCGGCGCTGTGTTCACTGCTGCCAGTGATGCAAGTGAACAGGCGAAGACGGCT 3163
QY 104 AlaLysAlaGluLysLeuAspArgLeuAlaThrThrThr-----SerGly 118
Db 3164 GCAAGTCCGATGTGTGGCAAGCAGTGGCGCCAGCACGGCTAACAAGTAATGCGAGTGCC 3223
QY 119 AlaLeuArgAlaThrPro-----PheAlaMetAlaSer 129
Db 3224 GCAGCCAGTGCAGACCAAGGCTGTGATAGCAAAAGCCGACAGCAGATTCGAGTGCAGCG 3283
QY 130 -LeuLeuGlnTyrMetGlnProAlaIleAsnLysGlyAspTrpLeuProAla----- 146
Db 3284 AGTGTGCAGCAAGCAGTGCAGAGGTGCAGAG-----CAGTTGCCACGGAAGCGGCG 3337
QY 147 ----ProLeuLysProLeuThrProLeuIleSerGlyAlaLeuSerGlyAla-----Me 163
Db 147 ----ProLeuLysProLeuThrProLeuIleSerGlyAlaLeuSerGlyAla-----Me 163

Db 3338 AGTGCCGGGCATCCGATGACTCG-CTAGCTTCTAGTGCCGCCAGTGCAGGCTT 3396
QY 163 tAspGlnValGlyThrLysMetMetAspArgAlaThrGlyAspLeuHisTyrLeuSerAl 183
Db 3397 TGACAAAGCTGCCAGCGCTGCGGAAGCGCAGCTCA-----AGTGC 3438
QY 183 aserProAspArgLeuHisAspAlaMetAlaAlaSerValLysArgHisSerProSerLe 203
Db 3439 CGCGAGCGGCGCTGCTAGTTCAGCGGCAGCTCAAGGCACACAGAGTGGCGCAAGCTCCAG 3498
QY 203 uAlaArgGlnValLeuAspThrGlyValAlaValGlnThrTyr----- 217
Db 3499 TGCCAGCGAAGCGGGTCAAGCATCAACCCGACATCTGTGTATGCTAGTGTCAAGTTC 3558
QY 218 -----SerAlaArgAsnAlaVal 223
Db 3559 GAGTGCAGTGAAGCCGGTTCATATGCTCATCAGGACGAGCTCCAGTGCAGTGAAGCGAC 3618
QY 223 l-----ArgThrValLeuAlaProAlaLeuAlaSerArgPr 235
Db 3619 TGGCCATGCAAGTAGTGCTACAAGTCAAGCAAGTAGTCCGCATCCAGTGTGCTCCAGGTA 3678
QY 235 oAlaValGlnGlyAlaValAspLeuGlyValSerMetAlaGlyGlyLeuAlaAlaAsnAl 255
Db 3679 CCCAAGTGTAGTGGGATCCAGTCCAGTCAATGTATTCGTCAGTGCACAGCAAGTACTGC 3738
QY 255 aGlyPheGlyAsnArgLeuLeuSerValGlnSerArgAspHisGlnArgGlyAlaLe 275
Db 3739 ATCC-----AGTCCCGCTAGTGCAGCCGACAAAGTGAGGCTTCGACGCG 3780
QY 275 uValLeuGlyLeuLysAspLysGluProLysAlaGlnLeuSerGluGluAsnAspTrpLe 295
Db 3781 ATCGTGGCTGCAAGTCAATGCTAGTGAACAGCAAGTATGCTTCCAGTGAAGATGTGT 3840
QY 295 uGluAlaTyrLysAlaIleLysSerAlaSerTyrSerGlyAlaAlaLeuAsnAlaGlyLy 315
Db 3841 ATCAAGCAGTGTGCTGCAGTGTGCTCCAGCGCGCCAGTGCAGTCCAGTGTGCAAA 3900
QY 315 s-----ArgMetAlaGlyLeuProLeuAspMetAlaThrAspAlaMetGlyAl 331
Db 3901 GGCTGTTAACAGTAGTGTGCGGGTATATACTCTCATGACGACAGTAGTGCACAGCAG 3960
QY 331 aValArgSerLeuValSerAlaSerSerLeuThrGlnAsnGlyLeuAlaLeuAlaGlyGl 351
Db 3961 TGCTTAAGACGGCTGAAGTCAAGCAAGCAGTGCAGCGCAGTGTGCTGCT----- 4009
QY 351 yPheAlaGlyValGlyLysLeuGlnLumetaLathrLysAsnIleThrAspProAlaTh 371
Db 4010 -----TCTGATGATTGCTGTAAGTCTTAGCGCTGC 4038
QY 371 rLysAlaAlaValSerGlnLeuThrAsnLeuAlaGlySerAlaAlaValPheAlaGlyTr 391
Db 4039 CAGTGCCTGCTTGTCTGACGATGCTAAGCAAGTAGCGCGCGATGTAGCATCCAGCGC 4098
QY 391 pThrThrAlaAlaLeuThrThrAspProAlaVal-----LysLy 404
Db 4099 TACAAGTGTGCCATTAAGTCCGACCATCTTGGCTGATCAGAGTGCACAGGGTCAAC 4158
QY 404 sAlaGluSerPheIleGlnAspThrValLysSerThrAlaSerSerThrThrGlyTyrVa 424
Db 4159 CGCTGGCTCCCATATT-----TTGCCAAGTACTGTGGAGAGACGACAGGATAGTAT 4209
QY 424 l---AlaAspGlnThr-----ValLysLeuAlaLysThrVa 435
Db 4210 ACCATCGGGTCAAGACGCCAACACAGCAAGCAAGCAACAGCAAGCAACAGCAAGCA 4269
QY 435 lLysAspMetGlyGlyGluAlaIleThrHisThrGlyAlaSerLeuArgAsnThrValAs 455
Db 4270 GCCAACACAAGCCGGTCAACACA---ACCAGACAGGT---TCATTACCGCAACAGGATCA 4323
QY 455 nAsnLeuArgGlnArgProAlaArgGluAlaAspIleGluGluGlyThrAla 473
Db 4324 TGCAGGCGAGCATATGCTACCGCACAGACCGGTGATGATGCTGAAGCGGTACTTCT 4378

RESULT 8	
AAQ29471	
ID	AAQ29471 standard; DNA; 6744 BP.
AC	AAQ29471;
XX	
DT	04-MAR-1993 (first entry)
XX	
DE	Extracellular factor related protein gene.
XX	
KW	EF*; detection; prevention; screening; diagnostic; ss.
OS	Streptococcus suis type II (non-pathogenic).
XX	
FH	Key
FT	CDS
FT	Location/Qualifiers
FT	361..5826
FT	/*tag= a
FT	361..498
FT	/*tag= b
FT	499..5826
FT	/*tag= c
FT	66..71
FT	/*tag= d
FT	/note= "-35 region"
FT	89..94
FT	/*tag= e
FT	/note= "-10 region"
FT	153..158
FT	/*tag= f
FT	/note= "-35 region"
FT	176..181
FT	/*tag= g
FT	/note= "-10 region"
FT	350..356
FT	/*tag= h
FT	/note= "ribosome binding site"
FT	2869
FT	/*tag= i
FT	/note= "start of repetitive unit R1"
FT	3097
FT	/*tag= j
FT	/note= "start of repetitive unit R2"
FT	3292
FT	/*tag= k
FT	/note= "start of repetitive unit R3"
FT	3520
FT	/*tag= l
FT	/note= "start of repetitive unit R4"
FT	4087
FT	/*tag= m
FT	/note= "start of repetitive unit R5"
FT	4381
FT	/*tag= n
FT	/note= "start of repetitive unit R6"
FT	4609
FT	/*tag= o
FT	/note= "start of repetitive unit R7"
FT	4837
FT	/*tag= p
FT	/note= "start of repetitive unit R8"
FT	5065
FT	/*tag= q
FT	/note= "start of repetitive unit R9"
FT	5293
FT	/*tag= r
FT	/note= "start of repetitive unit R10"
FT	5521
FT	/*tag= s
FT	/note= "start of repetitive unit R11"
FT	6554..6566
FT	/*tag= t
FT	/note= "dyad symmetry region 1a"

FT	misc_feature	6571..6583
FT	/tag= u	
FT	/note= "dyad symmetry region 1b"	
FT	6611..6625	
FT	/tag= v	
FT	/note= "dyad symmetry region 2a"	
FT	6631..6644	
FT	/tag= w	
FT	/note= "dyad symmetry region 2b"	
XX		
PN	WO9216630-A.	
XX		
PD	01-OCT-1992.	
XX		
Pf	19-MAR-1992;	92WO-NL00054.
XX		
PR	21-MAR-1991;	91NL-0000510.
XX		
PA	(DIER-) CENT DIERGENESKUNDIG INST.	
XX		
PI	Smith HE, Vecht U;	
XX		
DR	WPI; 1992-349215/42.	
DR	P-PsDB; AAR27745.	
XX		
PT	Deoxyribonucleic acid encoding virulence characteristic of	
PT	Streptococcus suis - useful for antibody and polypeptide for	
PT	diagnosing and preventing infections in pigs and humans	
XX		
PS	Claim 2; Fig 1b; 86pp; English.	
XX		
CC	The sequence is that of the extracellular factor related protein gene	
CC	from Streptococcus suis type II (non-pathogenic) which allows the	
CC	detection and the prevention of infections by S. suis in a more	
CC	effective manner than was previously possible. It facilitates screenin	
CC	of e.g. pigs and elimination of infected and carrier pigs can then be	
CC	carried out. The new diagnostic tests can distinguish between	
CC	avirulent and virulent strains. It may be used in the prodn. of	
CC	a vaccine. See also AAQ29470 and AAQ29472.	
XX		
SQ	Sequence 6744 BP; 2402 A; 1231 C; 1535 G; 1576 T; 0 other;	
Alignment Scores:		
Pred. No :	0.569	Length: 6744
Score:	144.50	Matches: 109
Percent Similarity:	37.04%	Conservative: 81
Best Local Similarity:	21.25%	Mismatches: 208
Query Match:	5.99%	Indels: 115
DB:	13	Gaps: 22
US-09-825-414-7 (1-486) x AAQ29471 (1-6744)		
QY	14 AlaThrAspSerPheArgThrAlaSerAspAlaSerLeuAlaSerSerValArgSer	33
	::: ::::: :::	
Db	4024 GCTAGAGATGCAGTTGAATTGGCTTAAGGATAAAGAATTAGGTAAAGAACATCCGAACA	4083
QY	34 ValSerSeraSpGlnGlnArgGluIleAsnAlaIleAlaasp-----	47
	::: ::::: ::::: ::::: ::::: :::::	
Db	4084 CAAGAAGAAGACCTACTTAAATAAGTAGAGAAACCTTGCAAGAAAGATACGCCCAAACGTATC	4143
QY	48 -----	TyrLeuThrAsp 51
Db	4144 GAGGACAATCCAAACTTGTCAGATGAAGATTAAGCAACGGAAATTAAAAAGCTTAACGTGAC	4203
QY	52 HisValPheAlaAlaHisLysLeuProProAlaAspSeraAlaasp-----	66
	:::	
Db	4204 GCTGTG---GCAAAAACCTTTAGCAACCATGCGTGAACAATGCAGATAAGCGTAGCGAAGAA	4260
QY	67 -----GlyGlnAlaAlaValAlaSpValHisAsnAlaGlnIleThrAlaLeuIleGlu	83
	::: :::	
Db	4261 GCAGAAAAAGCTCAAGCCCTAGCAGATCTTGAAAAAGCTAAAGAAACACAGAAAAATTGCA	4320
QY	84 ThrArgAlaSer-----ArgLeuHisPhe-----GluGlyGluThrProAlaThr	98

Db 4321 GATAAAGCTCGATGTGATGTTGACTATACTTGTGAAGATGGTGAGCTTGAAGCTACT 4380
:::|||||
QY 99 IleAlaAspThrPheAlaLysAlaGluLysLeuAspArgLeuAlaThrThrThrSergly 118
||| ||||| |||::: |||
Db 4381 AAACAAGAT-----GCTAAGACC--AAGATTGCTAAGATGCAGCCGCTGCTAAGAA 4431
QY 119 AlaLeuArgAlaThrProPheAlaMetAlaSerLeuLeuGlnTyrMetGlnProAlaIle 138
|||::: |||
Db 4432 CCAATTGCACGAATCCA----- 4449
QY 139 AsnLysGlyAspTrpLeuProAlaProLeuLysProLeuThrProLeuIleSerglyAla 158
||| ||| ||| ||| |||::: |||
Db 4450 AACTTGACAGAC-----GCAGAGAAGAAAACCTTCACCGATGCGTAGATGCAGAA 4500
QY 159 LeuSerglyAlaMetaspGlnValGlyThrLysMetMetLaspArgAlaThrGlyAspLeu 178
::: ||| |||::: |||
Db 4501 GTAGCCAAAGCTAAGCAGCGCAATTTCAGCT-----GCAACCAGC----- 4539
QY 179 HistyrLeuSeralAserProAspArgLeuHisAspAlaMetAlaAlaSerValLysArg 198
||| |||::: ||| ||| |||::: |||
Db 4540 -----CCAGCAGATGTTCAAAAAGAGAGAGATGCAGGTGTGCAGCCATTGCAGAA 4590
QY 199 HisSerProSerLeuAlaArgGlnValLeuAspThrGlyValAlaValGlnThrTyrSer 218
|||::: ||| |||::: |||
Db 4591 GATGTTCTTGACGCAGCTAAACAAGATGCTAAGAATAGATTGCTAAAGATGCAGCCGCT 4650
QY 219 AlaArgAsnAlaValArgThrValLeuAlaProAlaLeuAlaSerArgProAlaValGln 238
|||::: |||::: ||| |||::: |||
Db 4651 GCTAAAGAACCAATTGGCTCCAATCCAACTTGACAGACGACAGAGAAGAAAACCTTCACC 4710
QY 239 GlyAlaValAspLeuGlyValSerMetAlaGlyLysLeuAlaAlaAsnAlaGlyPheGly 258
||| ||| |||::: |||
Db 4711 GATCGCGTAGATGCAGAGTAGCCAAAGCTAACGACGCCAATTTCAGCTGCA----- 4761
QY 259 AsnArgLeuLeuSerValGlnSerArgAspHisGlnArgGlyAlaLeu----- 275
||| ||| |||::: |||
Db 4762 ACCAGCCCGACAGATGTTCAAAAAGAGAGCATGCAGGTGTTGCAGCCATTGCAGAGAT 4821
QY 276 ValLeuGlyLeuLysAspLysGluProLysAlaGlnLeuSergLysGluLysAspTrpLeu 295
||| |||::: ||| |||::: |||
Db 4822 GTTCTTGACGCAGCTAAACAAGATGCTAAGAATAGATTGCTAAGAATCCGAC----- 4875
QY 296 GluAlaTyrLysAlaIleLysSeralA-----SerTyrSerglyAlaAla 310
||| ||| ||| ||| |||::: |||
Db 4876 -----GCTGCTAAGTCAAGCCATTGACGCGCAATCCAACTTGACAGATGCAGAG 4923
QY 311 LeuAsnAlaGlyLysArgMetAlaGlyLeuProLeuAspMetAlaThrAspAlaMetGly 330
::: |||::: ||| ||| ||| |||::: |||
Db 4924 AAGGAATCAGCTAAGAAACAGTTGATGCTGATGCTAAAGCTGCGACAGATGCAAATTGAT 4983
QY 331 AlaValArgSerLeuValSerAlaSerSerLeuThrGlnAsnGlyLeuAlaLeuAlaGly 350
||| ||| ||| ||| |||
Db 4984 GCTTCAACAAGTCCAGTCGAAGCGCAATCGCAGAGACAAA----- 5025
QY 351 GlyPheAlaGlyValGlyLysLeuGlnGluMetAlaThrLysAsnIleThrAspProAla 370
||| ||| |||::: ||| |||::: |||
Db 5026 -----GGCGTAGGTTCAATC-----GCCCAAGATGTTCTTGACGACGCG 5064
QY 371 ThrLysAlaAlaValSerGlnLeuThrAsnLeuAlaGlySerAlaAlaValPheAlaGly 390
::: ||| |||::: ||| |||::: |||
Db 5065 AAACAAGATGCTAAGACACAGATTGCCAAGAAGTGGCCGACGCT----- 5109
QY 391 TrpThrThrAlaAlaLeuThrThrAspProAlaValLysLysAlaGlu--SerPheIle 409
|||::: |||::: ||| ||| ||| |||
Db 5110 -----AAGAAGCAATTGATGCCAATCCGAACCTTATCAGATGCAGAGAAGGAAGCTTCT 5163
QY 410 GlnAspThrValLysSerThrAlaSerSerThrThrGlyTyrValAlaAspGlnThrVal 429
::: ||| |||::: ||| ||| |||::: |||
Db 5164 AAGAAAGCGTAGATGCAGATGCTAAAGCTACGACAGATGCAATTGATGCTTCAACAAGT 5223
QY 430 LysLeu--AlaLysThrValLysAspMetGlyGluAlaIleThrHisThrGlyAla 448
::: |||::: ||| |||::: |||

Db 5224 CCAGTCGAAGCGCAATCGCGCAGAGACAAAGGC-----GTAAGTTCA 5265
QY 449 SerLeuArgAsnThrValAsnAsnLeuArgGln-----ArgProAlaArgGlu 464
::: |||::: ||| |||::: |||
Db 5266 ATGCCCAAGATGTTCTTGACGCAGCGAAACAAGATGCTAAGAATAGATTGCTAAGAA 5325
QY 465 AlaAspIleGluGluGlyGlyThrAlaAlaSerProSer 477
::: ||| |||::: ||| |||::: |||
Db 5326 TCCGACGCTGCTAGTCAAGCCATTGCAGCGCAATCCAAC 5364
RESULT 9
AAA10263
ID AAA10263 standard; DNA; 11883 BP.
XX
AC AAA10263;
XX
DT 03-JUL-2000 (first entry)
XX
DE Bordetella pertussis filamentous haemagglutinin gene, fhaB.
XX
KW Filamentous haemagglutinin; FHA; fhaB gene; surface-associated;
KW cellular adhesion; antigenic; immunogenic; whooping cough; pertussis;
KW diagnosis; prophylaxis; vaccine; ds.
XX
OS Bordetella pertussis.
XX
FH Key Location/Qualifiers
FT CDS 253..11043
FT /*tag= a
FT /product= "B. pertussis filamentous haemagglutinin"
XX
PN US6036960-A.
XX
PD 14-MAR-2000.
XX
PE 01-SEP-1994; 94US-0299941.
XX
PR 13-NOV-1989; 89US-0436297.
PR 10-AUG-1992; 92US-0928619.
PR 27-OCT-1988; 88US-0263648.
XX
PA (RELM/) RELMAN D A.
PA (DOME/) DOMENIGHINI M.
PA (RAPP/) RAPPUOLI R.
PA (FALK/) FALKOW S.
XX
PI Falkow S, Domenighini M, Relman DA, Rappuoli R;
XX
DR WPI; 2000-269872/23.
DR P-PSDB; AAY87407.
XX
PT New isolated nucleic acid encoding the filamentous hemagglutinin of
PT Bordetella pertussis, useful for diagnosis, treatment and prevention of
PT whooping cough -
XX
PS Claim 3; Columns 11-28; 17pp; English.
XX
CC This sequence represents the filamentous haemagglutinin gene, fhaB, from
CC Bordetella pertussis. Bordetella pertussis is a small Gram negative
CC bacillus found only in humans and which causes whooping cough (pertussis)
CC in children. Filamentous haemagglutinin (FHA) is a 368 kD surface-
CC associated protein which is thought to be one of the most important
CC factors in mediating the adhesion of the bacterium to a eukaryotic
CC (human) cell. FHA also stimulates an immune response in humans following
CC disease, and has been shown to act as an immunoprotective antigen in an
CC animal model. The FHA gene, or its expression products, are used for
CC prevention or treatment of pertussis, particularly in vaccines. It may
CC also be used to increase the amount of FHA in live or dead Bordetella
CC pertussis or other organisms. Fragments of the gene may be used as
CC diagnostic hybridisation probes or as antisense modulators to reduce the
CC infectivity of Bordetella pertussis. FHA, or fragments thereof, may be
CC used in immunoassays to detect anti-FHA antibodies, and to raise
CC antibodies for the diagnosis, prevention of treatment of whooping cough.

XX Sequence 11883 BP, 2196 A; 3648 C; 4282 G; 1754 T; 3 other;
SQ Alignment Scores:
Pred. No.: 1.31 Length: 11883
Score: 143.50 Matches: 112
Percent Similarity: 39.43% Conservative: 80
Best Local Similarity: 23.00% Mismatches: 205
Query Match: 5.95% Indels: 91
DB: 21 Gaps: 21
US-09-825-414-7 (1-486) x AAA10263 (1-11883)
QY 33 SerValSerSerAspGlnGlnArgGluIleAsnAlaIleAlaAspTyrLeuThrAspHis 52
DB 1492 GCGCGGAGCAGCCGTCAGCCCTGTCCTGAACGCCGGCGCCCTCAAGCGGACAAG 1551
QY 53 ValPheAlaAlaHisLysLeuProProAlaAspSerAlaAspGlyGlnAlaAlaValAsp 72
DB 1552 CTGTGGCGCAGCGCGGCTC-----GACGTGGATGGCAAGCAGCGCTCGCG 1599
QY 73 ValHisAsnAlaGlnIleThrAlaLeuIleGluThrArgAla----- 86
DB 1600 CTGGGTCGCGCAGCAGCAATGCGCTG--TCGGTCGCTGCCGGCGCGCCCTCAAGCG 1656
QY 87 -----SerArgLeuHisPheGluGlyGluThrProAlaThrIleAla 100
DB 1657 GGCAAGCTGTCCGGCAGCGGGCGCACTGGACGTGGACGGCAAGCAGCGCTCACGCTGGGT 1716
QY 101 AspThrPheAlaLysAlaGluLysLeuAspArgLeuAlaThrThrThrSerGlyAlaLeu 120
DB 1717 TCGGTTCCGAGCGACGGT-----GCCCTGTCTGGTAAGCGCTGGCGGAACCTG 1764
QY 121 ArgAlaThrProPheAlaMetAlaSerLeuLeuGlnTyrMet--GlnProAlaIleAsn 139
DB 1765 CGGGCGAACAATGTCTCTCAAGTCCCAACTTGAGTGCCTGGGCGAGCGGAGGTCCGG 1824
QY 140 LysGlyAspTyrLeuProAlaProLeuLysProLeuThrProLeuIleSerGlyAlaLeu 159
DB 1825 CTGGATGAC-----GCTTCAGCGCAGCGCGCATGACCTGTGTGCCGAGGACGCTG 1878
QY 160 Ser-----GlyAlaMetAspGlnValGly----- 167
DB 1879 GCGGCCCGCAACCTGCACTCCAAGGCGCCCATCGCGGTACAGGCTGAGAGCGGCTCAGC 1938
QY 168 -----ThrLysMetMetAspArgAlaThrGlyAspLeu 178
DB 1939 GTGGCCACACGCGAAGCAGCGGGAATTGCCGCGCGCGCGCGCCAGGTGATCTG 1998
QY 179 HisTyrLeuSerAlaSer-ProAspArgLeuHisAspAlaMetAlaAlaSerValLysAr 198
DB 1999 CACGACCTGAGCGCAGCGCGCGCGGATATCTCGGCGGAGGGCGCGCTCAATATCGGC 2058
QY 198 gHisSerProSerLeuAlaArgGlnValLeuAspThrGlyValAlaValGlnThrTyrSe 218
DB 2059 CGTGGCGGAGCGATACGATGTGAAGGTCTCCGCGCACGGCGCCTTGTGATGATAGC 2118
QY 218 rAlaArgAsnAlaValArgThrValLeuAlaProAlaLeuAlaSerArgProAlaValGl 238
DB 2119 ATGACGGCCCTCGGTGCGATCGCGCTCCAGGCGAGCGGACGCTGTCCGCCAAGATATG 2178
QY 238 nGlyAlaValAspLeuGlyValSerMetAlaGlyGlyLeuAlaAlaAsnAlaGlyPheGl 258
DB 2179 CGCAGCCGTGC-----GCCGTCACCGTCAGCGCGCGCGCGCTCAACCTGGGC----- 2229
QY 258 yAsnArgLeuLeuSerValGlnSerArgAspHisGlnArgGly--GlyAlaLeuValLe 277
DB 2230 -----GATGTCCAGTCGGATGGGCGAGTCCGGCCACCAAGCGCGCGCCAT 2276
QY 277 uGlyLeuLysAspLysGluProLysAlaGlnLeuSer--GluGluAsnAspTyrLeuGl 296
DB 2277 GACGGTGCAGACGTCCGCGCTGCCGCGACCTTCCGCTGCAGGGCGGCGACGCGTTCGA 2336

QY 296 uAla-----TyrLysAlaIleLysSerAlaSerTyrSerGlyAla----AlaLeuAsnAl 313
DB 2337 GCGCGGTTCTCTGAATTCGCCCGGTGCCATGACCGTGAACGGCCCGCATGCGGACT 2396
QY 313 aGlyLysArgMetAlaGlyLeuProLeuAspMetAlaThrAspAlaMetGlyAlaValAr 333
DB 2397 GGATGGCGCGCACCGCGGGCGGCAATTGCGGGTTTCCAGCAGCAGCGGCGAGCTGCGTTGGG 2456
QY 333 gSerLeuValSerAlaSerSerLeuThrGlnAsnGlyLeuAlaLeuAlaGlyGlyPheAl 353
DB 2457 CAGTCTCGCGGCCAAGGGCAGCTGACGGTATCG-----GCCGCGCGCGCGC 2504
QY 353 aGlyValGlyLysLeuGlnLeuMetAlaThrLysAsnIleThr----- 367
DB 2505 GACCGTGCGCGAGTTGAAGTCGCTGACACAACATCTCCGTGACGGCGCGGAACGCTGTC 2564
QY 368 -----AspProAlaThrLysAlaAlaValSerGlnLeuThrAsnLeu----- 381
DB 2565 GGTTCAGAGCGTCAACAGCGCGCTCCAGGTCGCCATTTCCGGCGCAGCGCGCTGGATGT 2624
QY 382 -----AlaGlySerAlaAlaValPheAlaGlyTyrThrAla----- 394
DB 2625 AGGCAAGTTTCCGCCAAGAGCGGTATCGGGCTCGAAGGCTGGGGCGCGCTCGGAGCGGA 2684
QY 395 -AlaLeuThrThrAspProAlaValLysLysAla-----GluSerPheIleGlnAspTh 412
DB 2685 CTCCCTCGGTTCCGACGGCGCGCATCAGCGTGTCCGGCGCGCATGCGGTGATGATCA 2744
QY 412 rValLysSerThrAlaSerSerThrThrGlyTyrValAlaAspGlnThr-----Va 429
DB 2745 AGCCCGCAGTCTTGGCGACATTTCTGCTGGGGCGGAAGCGCGCGCAGCTGGCGCGGT 2804
QY 429 lLysLeuAlaLysThrValLysAspMetGlyGlyGluAlaIleThr-----Hi 445
DB 2805 GGAGGCGCGCGGTTCGATCGACGTGCGCGCGGATCCACGCTGGCGGGAACCTCGCTGA 2864
QY 445 sThrGlyAlaSerLeuArg 451
DB 2865 CGCCAATCGCGACGTTCCG 2883
RESULT 10
AAZ20203
ID AAZ20203 standard; DNA; 1801 BP.
XX AC AAZ20203;
XX DT 17-JAN-2000 (first entry)
XX DE Mycobacterium tuberculosis antigen fusion protein Mt61f DNA.
XX KW Tuberculosis; antigen; fusion protein; Mt61f; TbH9; DPV; MTI;
XX KW diagnosis; therapy; vaccine; immunogen; ss.
XX OS Mycobacterium tuberculosis.
XX PN WO9951748-A2.
XX PD 14-OCT-1999.
XX PF 07-APR-1999; 99WO-US07717.
XX PR 07-APR-1998; 98US-0056556.
XX PR 30-DEC-1998; 98US-0223040.
XX PA (CORI-) CORIXA CORP.
XX PI Skeiky YAW, Alderson M, Campos-Neto A;
XX DR WPT; 1999-601610/51.
XX DR P-FSDB; AAY32068.
XX PT New fusion proteins useful for diagnosis, prevention and treatment of tuberculosis -

XX Example: Fig 10A-B; 83pp; English.
PS
XX This DNA sequence includes a coding region for a recombinant
CC Mycobacterium tuberculosis tri-antigen fusion protein (see AAY32068),
CC termed Mtb6lf, composed of the antigens Tbh9, DPV and MTI. The
CC DNA is useful for the recombinant production of the fusion protein.
CC Coding sequences for the antigens were modified by PCR in order
CC to facilitate their fusion and subsequent expression of the fusion
CC protein. Coding sequences for Tbh9, DPV and MTI were then ligated
CC to encode Mtb6lf. The invention provides fusion proteins (see
CC AAY32059-71) containing at least 2 M. tuberculosis antigens. The new
CC fusion proteins and polynucleotides encoding them are useful as
CC vaccines for preventing tuberculosis (claimed), for diagnosis (via
CC in vitro assays or intradermal skin tests for detection of anti-M.
CC tuberculosis antibodies), monitoring of disease progression, and
CC treatment of tuberculosis. They are more effective immunogens than
CC mixtures of the individual protein components.

XX Sequence 1801 BP; 288 A; 558 C; 642 G; 313 T; 0 other;

Alignment Scores:

Pred. No.:	0.157	Length:	1801
Score:	143.00	Matches:	117
Percent Similarity:	34.45%	Conservative:	58
Best Local Similarity:	23.03%	Mismatches:	165
Query Match:	5.93%	Indels:	168
DB:	20	Gaps:	25

US-09-825-414-7 (1-486) x AAZ20203 (1-1801)

QY 20 ThrAlaSerAspAlaSerLeuAlaSerSerSerValArgSerValSerSerAspGln 39
Db 262 ACCCGGGGAGGCGGAGCTGACCGCGCCAGTCCGGCTT----- 303
QY 40 ArgGluLeuAsnAlaIleAlaAspTyrLeuThrAspHisValPheAlaAlaHisLysLeu 59
Db 304 -----GTCGGCGCGCTTACGAGACGGCGGTATGGCTGACG-----GTG 342
QY 60 ProProAlaAspSerAlaAspGlyGlnAla----- 69
Db 343 CCCCCGCGGTGATCGCCGAGAACCCTGCTGAACGTGATCTGTATAGCAGCAACCTC 402
QY 70 -----AlaValAspValHisAsnAlaGlnIleThrAlaLeuIleGlu 83
Db 403 TTGGGGCAAAACACCCCCGGCGATCGCGGTCAACGAGGCCGAATACGGCGAGATGTGGCC 462
QY 84 ThrArgAlaSerArgLeuHisPheGluGluGluThrProAlaThrIleAlaAspThrPhe 103
Db 463 CAAGACGCGCGCGGATGTTGGCTACGCCGGCGGACGGCGGACGGCGGCGAGCTTG 522
QY 104 AlaLysAlaGluLysLeuAspArgLeuAlaThrThrThrSerGlyAlaLeu----- 120
Db 523 CTCGGCTTCGAGAGGCGCGGAGATG-----ACCAGCGCGGTGGCTCTCGAGCAG 576
QY 121 -----ArgAlaThrProPheAlaMetAlaSer---LeuLeuGlnTyrMet 134
Db 577 GCCCGCGCGGTGAGAGCGCTCCGACACCGCGCGGCAACCACTTGATGAACAAATGTG 636
QY 135 GlnProAlaIleAsnLysGlyAspTyrLeuProAlaProLeuLysProLeuThrProLeu 154
Db 637 CCCCAGCGCTGCAACAG-----CTGGCCACGCCACGACGAGGACACGACGCT--- 684
QY 155 IleSerGlyAlaLeuSerGlyAlaMetAspGlnValGly-----Thr 168
Db 685 ---TCTTCCAGCTGGGTGGCTGTGGAGAGCGGTCTGCCCATCGGTCCCGCATCAGC 741
QY 169 LysMetMetAspArgAlaThrGlyAspLeuHisTyrLeuSerAlaSerProAspArgLeu 188
Db 742 AACATGGTGTGATGGCCACACACACATGTG--ATGACCAACTCGGGTGTGTGATG 798
QY 189 HisAspAlaMetAlaAlaSerValLysArgHisSerProSerLeuAlaArgGlnValLeu 208
Db 189 HisAspAlaMetAlaAlaSerValLysArgHisSerProSerLeuAlaArgGlnValLeu 208

Db 799 ACCAACACCTTGAGCTCGATGTTGAAGGCTTTGCTCCGGCGGCGGCCAG----- 852
QY 209 AspThrGlyValAlaValGlnThrTyrSerAlaArgAsnAlaValArgThrValLeuAla 228
Db 853 -----GCCGTGCAAACC---GCCGCCAAACGGGGTCCGG----- 885
QY 229 ProAlaLeuAlaSerArgProAlaValGlnGlyAlaValAspLeuGlyValSerMetAla 248
Db 886 ---GCGATGAGCTCGCTGGGCGAGCTCGCTGGGTCTTCGGGCTCTGGGC----- 930
QY 249 GlyGlyLeuAlaAlaAsnAlaGlyPheGlyAsnArgLeuLeuSerValGlnSerArgAsp 268
Db 931 GGTGGGGTGGCCGCCCAACTTGGGT-----CGGGCGGCTCGGTGCGTTCG----- 975
QY 269 HisGlnArgGlyGlyAlaLeuValLeuGlyLeuLysAspLysGluProLysAlaGlnLeu 288
Db 976 -----TTG 978
QY 289 SerGluGluAsnAspTyrPheGluAlaTyrLysAlaIleLysSerAlaSer----- 305
Db 979 TCGGTGCCGAGCGCTGGCGCGCGCCCAACAGCAGCTACCCCGCGCGCGCGCTG 1038
QY 306 -----TyrSerGlyAlaAlaLeuAsnAlaGlyLysArgMetAlaGlyLeu 320
Db 1039 CCGGTGACCAGCCTGACACAGCGCCGGAAGAAGAGGCGCCGACAGATGCTGGCGGCGCTG 1098
QY 321 ProLeuAspMetAlaThrAspAlaMetGlyAlaValArgSerLeuValSerAlaSerSer 340
Db 1099 CCGGTGGG----- 1107
QY 341 LeuThrGlnAsnGlyLeuAlaLeuAlaGlyGlyPheAlaGlyVal----- 355
Db 1108 -----CAGATGGCGCGCAGCGCGGCTGTGGGTACAGTGTGTGCTGCTGCCGCCG 1161
QY 356 -----GlyLysLeu----- 358
Db 1162 CGACCTTATGTGATGCCGATCTCCGGCAGCGCGCAAGCTTGATCCCGTGACCGGCTC 1221
QY 359 -----GlnGluMetAlaThrLysAsnIleThrAspPro--- 369
Db 1222 ATTAACACACCTGCAATTAAGGGCAGGTAGTAGTGGCTCAACGCGAGATCCGGGG 1281
QY 370 -----AlaThrLysAlaAlaValSerGlnLeuThrAsnLeuAlaGly 383
Db 1282 GCTGCCGCACAGTTCACGCGCTCACCGGTGGCGCAGTCTTATTTGCGCAATTCCTCGCC 1341
QY 384 SerAlaAlaValPheAlaGlyTyrThrThrAlaAlaLeuThrThrAspProAlaValLys 403
Db 1342 GCACCGCCACCTCAGCGCGCTGCCATGGCCGCCCAATG-----CAAGCTGTGCCG 1392
QY 404 LysAlaGluSerPheIleGlnAspThrValLysSerThrAlaSerSerThrThrGlyTyr 423
Db 1393 GGGCGGGCACAGTACATC---GGCTTGTGAGTCCGTTGCCGCTCCTGCAACAACATAT 1449
QY 424 ValAlaAspGlnThrValLysLeuAlaLysThrValLysAspMetGlyGlyAlaIle 443
Db 1450 GAG-----CTCATGACGATTAATTACAGTTCGGGGAGCTGCAC 1488
QY 444 ThrHisThrGlyAlaSerLeuArg 451
Db 1489 GCTCAT---GGCGCATGATCCGC 1509
RESULT 11
AAK51943
ID AAK51943 standard; cDNA; 7545 BP.
XX AAK51943;
AC
XX
DT 06-NOV-2001 (first entry)
XX
DE Human polynucleotide SEQ ID NO 488.
XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW

KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KM tissue growth factor; immunomodulatory; cancer; leukaemia;
KW nervous system disorder; arthritis; inflammation; ss.
XX Homo sapiens.
XX WO200157190-A2.
XX
PD 09-AUG-2001.
XX
PF 05-FEB-2001; 2001WO-US04098.
XX
PR 03-FEB-2000; 2000US-0496914.
PR 27-APR-2000; 2000US-0560875.
PR 20-JUN-2000; 2000US-0598075.
PR 19-JUL-2000; 2000US-0620325.
PR 01-SEP-2000; 2000US-0654936.
PR 15-SEP-2000; 2000US-0663561.
PR 20-OCT-2000; 2000US-0693325.
PR 30-NOV-2000; 2000US-0728422.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drimanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
PI Xue AJ, Yang Y, Wejhrman T, Goodrich R;
XX
DR WPI; 2001-476283/51.
DR P-PSDB; AAM78810.
XX
PT Nucleic acids encoding polypeptides with cytokine-like activities,
PT useful in diagnosis and gene therapy -
XX
PS Claim 1; Page 1755-1761; 6221pp; English.
XX
CC The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC encoded polypeptides (AAM78323-AAAM80302) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activin/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
CC (AAM80020) are omitted as the relevant pages from the sequence listing
CC were missing at the time of publication.
XX
SQ Sequence 7545 BP; 1856 A; 2006 C; 1941 G; 1742 T; 0 other;

Alignment Scores:
Pred. No.: 0.912 Length: 7545
Score: 142.50 Matches: 125
Percent Similarity: 37.55% Conservative: 74
Best Local Similarity: 23.58% Mismatches: 250
Query Match: 5.91% Indels: 82
DB: 22 Gaps: 22

US-09-825-414-7 (1-486) x AAK51943 (1-7545)
QY 10 ProProValThrAlaThrAspSerPheArgThrAlaSerAspAlaSerLeuAlaSerSer 29
DB 885 CCATCGATGGCATCAACCGGTGCATCCGGGACATCGAGCAG-GCCTCGCTGGCGCCGCTC 943
QY 30 SerValArgSerValSerSerAspGlnGlnArgGluIleAsnAlaIleAlaAspTyrLeu 49
DB 944 AGC--CAGAGCCTGGCCACGAGGAGCAGACATCTGTGTGAGGCCCTGCAGAGCAGCTG 1000
QY 50 ThrAspHisValPheAlaIleAlaHisLysLeu-----ProProAlaAspSerAlaAspGly 67
DB 1001 ACTTCGGTGGTCCAGAAATCGGACACCTTATTCGATCCCATCGCCACAGCGGCTCGGGGA 1060

QY 68 GlnAlaAlaValAspValHisAsnAlaGlnIleThrAlaLeuIleGluThrArgAlaSer 87
DB 1061 GAAGCAGCTCAGCTGGACAT-----AAGGTGACACAACTGGCAAGC----- 1102
QY 88 ArgLeuHisPheGluGluThrProAlaThrIleAlaAspThrPheAlaLysAlaGlu 107
DB 1103 -----TATTTTGAG-----CCCTGATCTTAAGCCGACGTTGGTGGCTTCCAAAG 1147
QY 108 LysLeuAspArgLeuAlaThrThrThrSerGlyAlaLeuArgAlaThrProPheAlaMet 127
DB 1148 ATTCCTTGATCATCAGCAGCAGATGACG-----GTGCTGGACCAGACCAGACTCTCGCA 1201
QY 128 AlaSerLeuGlnTyrMetGlnProAlaIleAsnLysGlyAspTyrPheuProAlaPro 147
DB 1202 GAGTCTGCCTTGCAG---ATGTTGTATGCAGCCAAAGAGGTGGCGAAACCCCAAGGCA 1258
QY 148 LeuLysProLeuThrProLeuIleSerGlyAla-----LeuSerGlyAlaMetAspGln 165
DB 1259 CAACACACCCATGACGCCATCACAGAGGCCGCCAGTTGATGAAGAGAGCCGTGATGAC 1318
QY 166 ValGlyThrLysMetMetAspArgAlaThrGlyAspLeuHisTyrLeuSerAlaSerPro 185
DB 1319 ATC--ATGGTGACCGTGAACGGAAGCTGCCAGTGAAGTGGGGCTGGTGGGGCATGCTG 1375
QY 186 AspArgLeuHisAspAlaMetAlaAlaSerValLysArgHisSerProSerLeuAlaArg 205
DB 1376 GACGCCATTTGCAGAGCCATGAGCAAGCTGGATGAAGGCACCTCCTCCAGAACCAAGGGA 1435
QY 206 GlnValLeuAspThrGlyValAlaValGlnThrTyrSerAlaArgAsnAlaVal----- 223
DB 1436 ACATTTGTGCATTACAGACGACTGTGTTAAATACTCCAAAAGCATTTGGGTGACAGCT 1495
QY 224 -----ArgThrValLeuAlaProAlaLeuAlaSerArgProAlaValGln 238
DB 1496 CAGGAATGATGACTAAGTCGGTTACTTAACCCGAGAGAGTTGGAGAGACTGGCTTCACAA 1555
QY 239 GlyAlaValAspLeuGly---ValSerMetAlaGlyGlyLeuAlaAla----- 253
DB 1556 ATGACCAGTGACTATGGGCACCTGGCTTCCAGGCCAAGATGGCAGCAGCAGCGCGGA 1615
QY 254 -----AsnAlaGlyPheGlyAsnArgLeuLeuSerValGlnSerArgAspHis----- 269
DB 1616 CCAGAGGAGATCGGATTTCCAGATTCCG---ACTGCTGTGCAGAGACCTGGGCCACGGCTGT 1672
QY 270 -----GlnArgGlyGlyAlaLeuValLeuGlyLeuLysAspLysGluProLys 285
DB 1673 ATCTTCCTGGTGCAAGAGGAGGGGCCCTCCAGGCTGTGCCACAGACAGCTACACCAAG 1732
QY 286 AlaGlnLeuSerGlu---GluAsnAspTyrPheGluAlaTyrLysAlaIleLysSerAla 304
DB 1733 AGGAGCTGATCGAATGCGCCCGTGCCTACCGGAAAAGGTCTCTTGGTCTCTCGGCT 1792
QY 305 SerTyrSerGly-----AlaAlaLeuAsnAlaGlyLysArgMetAlaGly 319
DB 1793 CTCGAGCGCGGAACAAGAACCCAGGCATGATTACAGCCGACCGCTGTGTGGG 1852
QY 320 LeuProLeuAspMetAlaThrAspAlaMetGlyAlaValArgSerLeuValSerAlaSer 339
DB 1853 ATCATTTGCCGACCTGGACACACCATATTATGTTGCCAACACAGCGGGAGCGCTGAATGCAGAG 1912
QY 340 SerLeu-----ThrGlnAsnGlyLeuAlaLeuAlaGlyGlyPheAla 353
DB 1913 AACAGTGAGACCTTCCGACAGACACAGGAGAGAACATTCTCAAGACGGCCAGCCCTTGTA 1972
QY 354 GlyValGlyLysLeuGlnGluMetAlaThrLysAsnIleThrAspProAlaThrLysAla 373
DB 1973 GAAGACAGCAACTACTTGTGTACAGAGCTGGCTCCACTCTGACAAAGCTGGCCACAGCG 2032
QY 374 AlaValSerGlnLeuThrAsnLeuAlaGlySerAlaAlaValPheAlaGlyTyrThrThr 393
DB 2033 GCCCAGTCTCAGAGGCCACCATCACCCAGCTGCCAGAAAGTGTCAAGCTGGGGCAGGCC 2092

QY 394 AlaAlaLeuThrThrAspPro-----AlaValLysLysAla 405
::: |||||
Db 2093 AGCTGGGCTCCGACGACCCCGAGACCAGGTGGTTTGATCATGCCATCAAGATGTG 2152
QY 406 GluSerPheIleGlnAspThrValLysSerThrAlaSerSerThrThrGlyTyrValAla 425
Db 2153 GCCAAGGCCCTTTCTGATCTCATCATGCTACCAAGGAGCTGCCAAGCCAGTGGAC 2212
QY 426 AspGlnThrVal-----Lys 430
||| :::::
Db 2213 GACCCCTTCATGTACAGACTCAAGGGGCTGCCAAGGTGATGTTGACCAATGTCACTCG 2272
QY 431 LeuAlaLysThrValLysAspMetGlyGlyGluAlaIleThrHisThrGlyAlaSerLeu 450
||| ||||| ||||| :::
Db 2273 CTCCTCAAGACTGTAAAGCAGTGGAGGATGAGGCCACCCGGGACACGCGG---CTT 2329
QY 451 ArgAsnThrValAsnAsnLeuArgGlnArgProAla-----ArgGluAlaAspIleGlu 468
|||::: ::::: |||
Db 2330 GAGGCCACAATTGAATGCATAAAGCAGAGCTTACGGTGTTCAGTCAAAAGACGTACT 2389
QY 469 GluGlyGlyThrAlaAlaSerProSerGlu 478
||| ::||| |||
Db 2390 GAA-----AAGACATCATCACCTGAAGAA 2413
RESULT 12
AAI58449
ID AAI58449 standard; cDNA; 7615 BP.
XX
AC AAI58449;
XX
DT 22-OCT-2001 (first entry)
XX
DE Human polynucleotide SEQ ID NO 652.
XX
KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia; ss.
XX
OS Homo sapiens.
XX
PN WO200153312-A1.
PD 26-JUL-2001.
XX
PF 26-DEC-2000; 2000WO-US34263.
XX
PR 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-0552317.
PR 09-JUL-2000; 2000US-0598042.
PR 19-JUL-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0653450.
PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX
DR WPI; 2001-442253/47.
DR P-PSDB; AAM39293.
XX
PT Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
XX
PS Claim 1; SEQ ID NO 652; 10078bp; English.
XX
CC The invention relates to human nucleic acids (AAI57798-AAI61369) and

CC the encoded polypeptides (AAM38642-AAM42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX
SQ Sequence 7615 BP; 1874 A; 2026 C; 1954 G; 1754 T; 7 other;
Alignment Scores:
Pred. No.: 0.922 Length: 7615
Score: 142.50 Matches: 125
Percent Similarity: 37.55% Conservative: 74
Best Local Similarity: 23.58% Mismatches: 250
Query Match: 5.91% Indels: 82
DB: 22 Gaps: 22
US-09-825-414-7 (1-486) x AAI58449 (1-7615)
QY 10 PROProValThrAlaThrAspSerPheArgThrAlaSerAspAlaSerLeuAlaSerSer 29
||| ::||| ::||| |||::||| ||||| |||||:::
Db 955 CCATCGATGGCATCAACCGGTGCATCCGGGACATCGAGCAG-GCCTCGCTGGCCGCCGTC 1013
QY 30 SerValArgSerValSerSerAspGlnGlnArgGluIleAsnAlaIleAlaAspTyrLeu 49
||| ::|||::: |||::: |||::: |||::: |||
Db 1014 AGC---CAGAGCCTGGCCACGAGGAGCAGACATCTGTGGAGGCCCTGCAGAGCAGCTG 1070
QY 50 ThrAspHisValPheAlaAlaHisLysLeu-----ProProAlaAspSerAlaAspGly 67
||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1071 ACTTCGGTGTCCAGAAATCGGACACCTTATCGATGCCATCGCCACAGCGGCTCGGGGA 1130
QY 68 GlnAlaAlaValAspValHisAsnAlaGlnIleThrAlaLeuIleGluThrArgAlaSer 87
::: ||||| ||| :::: ||| |||
Db 1131 GAAGCAGCTCAGCTGGGACAT-----AAGTGACACACTGGCAAG----- 1172
QY 88 ArgLeuHisPheGluGlyGluThrProAlaThrIleAlaAspThrPheAlaLysAlaGlu 107
::: ||||| ||| ::||| ::||| ::||| ::|||
Db 1173 -----TATTTGAG-----CCCTTGATCTTAGCCCGCAGTTGGTGTGGCCCTCCAAG 1217
QY 108 LysLeuAspArgLeuAlaThrThrThrSerGlyAlaLeuArgAlaThrProPheAlaMet 127
||||| ||||| ||| ||| ||| ||| ||| ||| |||
Db 1218 ATTCTTGATCATCAGCAGCAGATGACG-----GTGTTGAGCCAGACCAAGACTCTCGCA 1271
QY 128 AlaSerLeuGlnTyrMetGlnProAlaIleAsnLysGlyAspTyrLeuProAlaPro 147
||| ||||| ||| ||| ||| ||| ||| ||| |||
Db 1272 GAGTCTGCCTTGCAG--ATGTTGTATGCAGCCAAAGAGGTGGCGGAACCCCAAGCA 1328
QY 148 LeuLysProLeuThrProLeuIleSerGlyAla-----LeuSerGlyAlaMetAspGln 165
::: ::||| ::||| ::||| ::||| ::|||
Db 1329 CAACACACCCATGACGCCATGCACAGAGGCCGCCAGTTGATGAAGAGGACCGGTGATGAC 1388
QY 166 ValGlyThrLysMetMetAspArgAlaThrGlyAspLeuHisTyrLeuSerAlaSerPro 185
::: ::||| ::||| ::||| ::||| ::|||
Db 1389 ATC---ATGGTGACGCTGAACGAAAGCTGCCAGTGAAGTGGGGCTGGTGGGGGCAATGCTG 1445
QY 186 AspArgLeuHisAspAlaMetAlaAlaSerValLysArgHisSerProSerLeuAlaArg 205
||| ::: ||||| ::||| ::||| |||
Db 1446 GAGGCCATTGGCAGAGCCATGAGCAAGCTGGATGAAGGCACTCTCCAGAACCAAGGGA 1505
QY 206 GlnValLeuAspThrGlyValAlaValGlnThrTyrSerAlaArgAsnAlaVal----- 223
::: ||| ::||| ||| ||| ||| ||| ||| ||| |||
Db 1506 ACATTTGTGACTATCAGACGAGCTGTGTTAAATACTCCAAAAGCATTTGGCGTGACAGCT 1565
QY 224 -----ArgThrValLeuAlaProAlaLeuAlaSerArgProAlaValGln 238

Db 1566 CAGAAATGACTAAGTCGGTTACTTAACCCGGAGAGTGGAGAGCTGGCTTCACAA 1625
QY 239 GYAlaValAspLeuGly--ValSerMetAlaGlyGlyLeuAlaAla----- 253
Db 1626 ATGACCAGTGACTATGGGCACCTGGCTTCCAGGGCCAGATGGCAGCACCACGGCGAA 1685
QY 254 -----AsnAlaGlyPheGlyAsnArgLeuLeuSerValGlnSerArgAspHis----- 269
Db 1686 CCAGAGGAGATCGGATTCAGATTCCG---ACTCGTGTGCAGGACCTGGGCCACGGCTGT 1742
QY 270 -----GlnArgGlyGlyAlaLeuValLeuGlyLeuLysAspLysGluProLys 285
Db 1743 ATCTTCTGTGTGCAGAAAGCAGGGGCCCTCCAGGTCCTGCCCCACAGACGCTACACCAAG 1802
QY 286 AlaGlnLeuSerGlu--GluAsnAspTrpLeuGlnAlaTyrLysAlaIleLysSerAla 304
Db 1803 AGGAGCTGATCGAATGCGCCCGTGCCTGCACGAAAGGTCCTCCTGGTGTCTCGGCT 1862
QY 305 SerTyrSerGly-----AlaAlaLeuAsnAlaGlyLysArgMetAlaGly 319
Db 1863 CTCACGGCCGGGAACAAGAACCCAGGCGATGCATTACAGCCGCCACCGCTGTCTGGG 1922
QY 320 LeuProLeuAspMetAlaThrAspAlaMetGlyAlaValArgSerLeuValSerAlaSer 339
Db 1923 ATCATTTGCCGACCTGGACACCACTATTATGTTTGCACACAGCGGGGACGCTGAATGCAGAG 1982
QY 340 SerLeu-----ThrGlnAsnGlyLeuAlaLeuAlaGlyGlyPheAla 353
Db 1983 AACAGTGAGACCTTCGCAGACACACAGGGAGAACATTCTCAAGACGGCCCAAGCCTTGTA 2042
QY 354 GlyValGlyLysLeuGlnGluMetAlaThrLysAsnIleThrAspProAlaThrLysAla 373
Db 2043 GAAGACACGAAACTACTTGTGTACAGAGCTGCGTCCACTCTGTACACAGCTGGCCACAGCG 2102
QY 374 AlaValSerGlnLeuThrAsnLeuAlaGlySerAlaAlaValPheAlaGlyTrpThrThr 393
Db 2103 GCCCAGTCTCTGACAGCCACCACCATCACCCAGCTCGCAGAAGTGTCAAGCTGGGGGCAGCC 2162
QY 394 AlaAlaLeuThrThrAspPro-----AlaValLysLysAla 405
Db 2163 AGCCTGGGCTTCGACGACGCCGAGACCCAGCTGTTTGTATCAATGCCATCAAAAGATGTG 2222
QY 406 GluSerPheIleGlnAspThrValLysSerThrAlaSerSerThrThrGlyTyrValAla 425
Db 2223 GCCAAGGCCCTTCTGATCTCATCATGCTTAACCAAGGAGCTGCCAGCAAGCCAGTGAC 2282
QY 426 AspGlnThrVal-----Lys 430
Db 2283 GACCCCTTCATGTACCAGCTCAAGGGGCTGCCAAGGTGATGCTGACCAATGTCACTCG 2342
QY 431 LeuAlaLysThrValLysAspMetGlyGlyAlaIleThrHisThrGlyAlaSerLeu 450
Db 2343 CTCCTCAAGACTGTAAAGGAGTGAGAGTGAAGCCACCCGGGGCACCAAGGGCG--CTT 2399
QY 451 ArgAsnThrValAsnAsnLeuArgGlnArgProAla-----ArgGlnAlaAspIleGlu 468
Db 2400 GAGGCCACAAATTGAATGCATAAAGCAGAGCTTACGGTGTTCAGTCAAAAGACGTACCT 2459
QY 469 GluGlyGlyThrAlaAlaSerProSerGlu 478
Db 2460 GAA-----AAGACATCATCACCTGAAGAA 2483
RESULT 13
AAI58468
ID AAI58468 standard; cDNA; 7657 BP.
XX
AC AAI58468;
XX
DT 22-OCT-2001 (first entry)
XX
DE Human polynucleotide SEQ ID NO 671.
XX

KW Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer;
KW peripheral nervous system; neuropathy; central nervous system; CNS;
KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;
KW amyotrophic lateral sclerosis; Shy-Drager Syndrome; chemotactic;
KW chemokinetic; thrombolytic; drug screening; arthritis; inflammation;
KW leukaemia; ss.
XX
OS Homo sapiens.
XX
PN WO200153312-A1.
XX
PD 26-JUL-2001.
XX
PF 26-DEC-2000; 2000WO-US34263.
XX
PR 21-JAN-2000; 2000US-0488725.
PR 25-APR-2000; 2000US-0552317.
PR 09-JUL-2000; 2000US-0598042.
PR 19-JUL-2000; 2000US-0620312.
PR 03-AUG-2000; 2000US-0653450.
PR 14-SEP-2000; 2000US-0662191.
PR 19-OCT-2000; 2000US-0693036.
PR 29-NOV-2000; 2000US-0727344.
XX
XX (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;
PI Wang J, Wang Z, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;
PI Zhao QA, Zhou P, Goodrich R, Drmanac RT;
XX
XX WPI: 2001-442253/47.
DR P-PsDB; AAM39312.
DR
XX
PT Novel nucleic acids and polypeptides, useful for treating disorders
PT such as central nervous system injuries -
XX
XX Claim 1; SEQ ID NO 671; 10078pp; English.
PS
XX
CC The invention relates to human nucleic acids (AAI57798-AAI61369) and
CC the encoded polypeptides (AAM38642-AA42213) with nootropic,
CC immunosuppressant and cytostatic activity. The polynucleotides are useful
CC in gene therapy. A composition containing a polypeptide or polynucleotide
CC of the invention may be used to treat diseases of the peripheral nervous
CC system, such as peripheral nervous injuries, peripheral neuropathy and
CC localised neuropathies and central nervous system diseases, such as
CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the
CC utilisation of the activities such as: Immune system suppression,
CC Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,
CC assays for receptor activity, arthritis and inflammation, leukaemias and
CC C.N.S disorders.
CC Note: The sequence data for this patent did not form part of the printed
CC specification.
XX
SQ Sequence 7657 BP; 1889 A; 2036 C; 1971 G; 1761 T; 0 other;
XX
Alignment Scores:
Pred. No.: 0.928 Length: 7657
Score: 142.50 Matches: 125
Percent Similarity: 37.55% Conserves: 74
Best Local Similarity: 23.58% Mismatches: 250
Query Match: 5.91% Indels: 82
DB: 22 Gaps: 22
US-09-825-414-7 (1-486) x AAI58468 (1-7657)
QY 10 ProProValThrAlaThrAspSerPheArgThrAlaSerAspAlaSerLeuAlaSerSer 29
Db 997 CCATGATGGCATCAACCGGTGCATCCGGGACATCGAGCAG-GCCTTCGCTGGCCGCCGTC 1055
QY 30 SerValArgSerValSerSerAspGlnArgGluIleAsnAlaIleAlaAspTyrLeu 49
III :::::::::::::: :::: ::::: |||

Db 1056 AGC---CAGAGCCTGGCCACGAGGAGCAGACATCTGTGGAGGCCCTGCAGAGCAGCTG 1112
QY 50 ThrAspHisValPheAlaAlaHisLysLeu-----ProProAlaAspSerAlaAspGly 67
Db 1113 ACTTCGGTGGTCCAGAAATCGGACACCTTATGCAATCCCATCGCCACAGCGGCTCGGGGA 1172
QY 68 GlnAlaAlaValAspValHisAsnAlaGlnIleThrAlaLeuIleGluThrArgAlaSer 87
Db 1173 GAAGCAGCTCAGCTGGGACAT-----AAGGTGACACAACCTGGCAAGC----- 1214
QY 88 ArgLeuHisPheGluGlyGluThrProAlaThrIleAlaAspThrPheAlaLysAlaGlu 107
Db 1215 -----TATTTTGAG-----CCCTTGATCTTAGCCCGCAGTTGGTGTGGCCTCCAG 1259
QY 108 LysLeuAspArgLeuAlaThrThrThrSerGlyAlaLeuArgAlaThrProPheAlaMet 127
Db 1260 ATTCTTGATCATCAGCAGCAGATGACG-----GTGCTGGACCAGACCAAGACTCTCGCA 1313
QY 128 AlaSerLeuLeuGlnTyMetGlnProAlaIleasnLysGlyAspTrpLeuProAlaPro 147
Db 1314 GAGTCTGCCTTGACAG---ATGTTGTATGACGCCAAAGAGGTGGCGAAACCCCAAGGCA 1370
QY 148 LeuLysProLeuThrProLeuIleSerGlyAla-----LeuSerGlyAlaMetAspGln 165
Db 1371 CAACACACACCATGACGCCCATCACAGAGGCCGCCAGTTGATGAAGAGAGCCGTGGATGAC 1430
QY 166 ValGlyThrLysMetMetAspArgAlaThrGlyAspLeuHisTyrLeuSerAlaSerPro 185
Db 1431 ATC---ATGGTGACGCTGAACGAAGCTGCCAGTGAAGTGGGGCTGGTGGGGGCATGGTG 1487
QY 186 AspArgLeuHisAspAlaMetAlaAlaSerValLysArgHisSerProSerLeuAlaArg 205
Db 1488 GACGCCCATTGACAGAACCATGAGCAAGCTGGATGAAGGCACCTCTCCAGAACCAAGGGA 1547
QY 206 GlnValLeuAspThrGlyValAlaValGlnThrTyrSerAlaArgAsnAlaVal----- 223
Db 1548 ACATTTTGTCAGACTATCAGACGACTGTGTTAAATACTCCAAAGCATTTGCCGTGACAGCT 1607
QY 224 -----ArgThrValLeuAlaProAlaLeuAlaSerArgProAlaValGln 238
Db 1608 CAGGAATGATGACTAAGTCGGTTACTAACC CGAGAGAGTTGGAGAGACTGGCTTCACAA 1667
QY 239 GlyAlaValAspLeuGly---ValSerMetAlaGlyGlyLeuAlaAla----- 253
Db 1668 ATGACCAAGTACTATGGCAGCTGGCTTCCAGGGCCAGATGGCAGCAGCCAGCGCGAA 1727
QY 254 -----AsnAlaGlyPheGlyAsnArgLeuLeuSerValGlnSerArgAspHis----- 269
Db 1728 CCAGAGGAGATCGGATTCAGATTCGC---ACTCGTGTCCAGGACCTGGGGCCACGGCTGT 1784
QY 270 -----GlnArgGlyGlyAlaLeuValLeuGlyLeuLysAspLysGluProLys 285
Db 1785 ATCTTCCTGGTGACAGAGCGGGCCCTCCAGGTCTGCCCCACAGACAGCTACACCAAG 1844
QY 286 AlaGlnLeuSerGlu---GluAsnAspTrpLeuGluAlaTyrLysAlaIleLysSerAla 304
Db 1845 AGGAGAGCTGATCGAATGGCGCCCTGCCGTACAGGAAAGGTCTCTGTGCTCTCGGCT 1904
QY 305 SerTyrSerGly-----AlaAlaLeuAsnAlaGlyLysArgMetAlaGly 319
Db 1905 CTCACAGCGCGGACAAGAACCCAGGCGATGCATTCACAGCCGCCACCGCTGTGTGGG 1964
QY 320 LeuProLeuAspMetAlaThrAspAlaMetGlyAlaValArgSerLeuValSerAlaSer 339
Db 1965 ATCATTTGCCGACCTGGACACCAACCATTTATGTTTGCACACGCGGGAGCGCTGAATGCAGAG 2024
QY 340 SerLeu-----ThrGlnAsnGlyLeuAlaLeuAlaGlyGlyPheAla 353
Db 2025 AACAGTGAGACCTTCGCAGACACACAGGAGAACATTTCAAGACGCGCAAGGCCCTTGCTA 2084
QY 354 GlyValGlyLysLeuGlnGluMetAlaThrLysAsnIleThrAspProAlaThrLysAla 373
Db 2085 GAAGACACGAAACTACTTGTGTCTCAGGAGCTGCGTCCACTCTGACAAAGCTGGCCCAAGCG 2144

QY 374 AlaValSerGlnLeuThrAsnLeuAlaGlySerAlaAlaValPheAlaGlyTrpThrThr 393
Db 2145 GCCCAGTCTTCAGCAGCCCAACCATCACCCAGCTCGCAGAAAGTGCTCAAGCTGGGGCAGCC 2204
QY 394 AlaAlaLeuThrThrAspPro-----AlaValLysLysAla 405
Db 2205 AGCCTGGGCTCCGACGACCCCCGAGACCAGGAGTGGTTTGATCAATGCCATCAAGATGTG 2264
QY 406 GluSerPheIleGlnAspThrValLysSerThrAlaSerSerThrThrGlyTyrValAla 425
Db 2265 GCCAAGGCCCTTTCTGATCTCATCTACGTCTACCAAGGAGCTGCCAGCAAGCAGTGGAC 2324
QY 426 AspGlnThrVal-----Lys 430
Db 2325 GACCCCTTCCATGTACCAGCTCAAGGGGCTGCCAAGTGTGATGTGACCAATGTCACTCG 2384
QY 431 LeuAlaLysThrValLysAspMetGlyGlyGluAlaIleThrHisThrGlyAlaSerLeu 450
Db 2385 CTCCTCAAGACTGTAAGGCAGTGGAGGATGAGGCCCAACCCGGGGCACCAGGGCG---CTT 2441
QY 451 ArgAsnThrValAsnAsnLeuArgGlnArgProAla-----ArgGluAlaAspIleGlu 468
Db 2442 GAGGCCACAATGTAATGCAATAAAGCAGAGCTTACGGTGTTCAGTCAAAAGACGTACCT 2501
QY 469 GluGlyGlyThrAlaAlaSerProSerGlu 478
Db 2502 GAA-----AAGACATCATCACCTGAAGAA 2525

RESULT 14
ABL27385
ID ABL27385 standard; DNA; 2466 BP.
XX
AC ABL27385;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 33628.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical; gene; ds.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Claim 1; SEQ ID NO 33628; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).

CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

xx Sequence 2466 BP; 659 A; 727 C; 740 G; 340 T; 0 other;

Alignment Scores:

Pred. No.:	0.318	Length:	2466
Score:	141.00	Matches:	129
Percent Similarity:	34.47%	Conservative:	74
Best Local Similarity:	21.90%	Mismatches:	198
Query Match:	5.85%	Indels:	189
DB:	23	Gaps:	25

US-09-825-414-7 (1-486) x ABL27385 (1-2466)

QY	2	HisIleAsnArgArgValGlnGlnProProValThrAlaThrAspSerPheArgThrAla	21
DB	549	CAGCCCATTCGCCGGGGAAGTCGCCACCACGACACACAATATGG---CGAGGCGGA	605
QY	22	SerAspAlaSerLeuAlaLaserSerSerValArgSerValSerSerAspGlnArgGlu	41
DB	606	ACAGAC-----CAGCGGCAA	620
QY	42	IleAsnAlaIleAlaAspTyrLeuThrAspHisValPheAlaAla-----HisLys	58
DB	621	CTTCTGCG---ACCCATTGCCCTTCAAGCCCATTTCCATTTGAGCCGGACTACCGCATCGC	677
QY	59	LeuProProAlaAspSerAlaAspGlyGlnAlaAlaValAspValHisAsnAlaGlnIle	78
DB	678	CTGCCAGCAGCAGCAGCAGCATCAGCAGCAACTGCAGTT-----GCAACAACA	725
QY	79	ThrAlaLeuIleGluThrArgAlaSerArgLeuHisPheGluGlyGluThrProAlaThr	98
DB	726	GCAGCAACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGTTCAGCTGCAACACCAACAGCA	785
QY	99	IleAlaAspThrPheAlaLysAlaGluLysLeuAspArgLeuAlaThrThrThrSerGly	118
DB	786	GCAGCAGCATCAACACCAACATCAACATCA-----ACATCAACAACAACATCAAGTC	836
QY	119	Ala-----Leu	120
DB	837	GCTGCAGCTGCCGACGAGGGGGGCGGCGGCGGCGCTGGCGGCCGAGCGGTACGGCTA	896
QY	121	Arg-AlaThrPro-----PheAlaMetAl	128
DB	897	CGGATCGACGGCCGTCACCTTGCCCCACTGCCACCGCATCGGCCAGAAATTGGCAGTAC	956
QY	128	AserLeuLeuGlnTyr-----MetGlnProAlaIle	138
DB	957	CACCTGACCTACGACACTTGGCCGCTGAGCAGCAATCATCGCCTGCTGCAGCGTTGCAG	1016
QY	138	eAsnLysGlyAspTyrLeuProAlaProLeu-----LysPro--	150
DB	1017	TAAAGAG---GACGCCCTGACCCCTTGACCTCTTGAGCAGCAGCGGCAGCCACAGACCCGA	1073
QY	151	----LeuThrProLeuIleSerGlyAlaLeuSerGlyAlaMetAsp-----	164
DB	1074	TGGTGCCAGCAGCAGCAGCAGCGGCACTGTGGTGGCAGCAGCGACTTGACGCCCTCGCC	1133
QY	165	-----GlnValGlyThrLysMetMetAspArgAlaThrGlyAspLe	178
DB	1134	GTCGGATTCTGGGATCTCGGAACCTGGAGGGCGCGCTGAAGATCGC---GATTCCGAGCT	1190
QY	178	uHisTyrLeuSerAlaSerProAspArgLeuHisAspAlaMetAlaAlaSerValLysAr	198
DB	1191	CTCGTACCTGGCAGAGCCATGAGCACAATGA-GAAAGACAAGAGGTGTACTGGGAGC	1249
QY	198	gHisSerProSerLeuAlaArgGlnValLeuAspThrGlyValAlaValGlnThrTyrSe	218
DB	1250	ACGAGACGACGCGTTGGCGCTCTTACGAGGGRCAAGCAGCGGAGTGTCAACTGAAGC	1309
QY	218	rAlaArgAsn-----AlaValArgThrValLeuAlaProAlaLeuAlaSerArgProAl	236

DB	1310	TGCGCAAGATGAGCAGCTACTCGGACTGCAGCAGCTTCCAGCTGAAGCAGCAGCAGCTGC	1369
QY	236	aValGlnGlyAla-----ValAspLe	243
DB	1370	GCCAATCGAGAGCGTCAATCGGCTGCAGCAGCAGCTCGATTGTGGCCAGAGACTCCAGCC	1429
QY	243	uGlyValSerMetAlaGlyGlyLeuAlaAlaAsnAlaGlyPheGlyAsnArgLeuLeuSe	263
DB	1430	AGGGGCTGCACACAGAGGTGGACCGCACTGCCCTCACAGCTGGAGAGCAGCGAGTGGCGG	1489
QY	263	rValGln-----SerArgAspHisGlnArgGlyGlyAla-----	274
DB	1490	TCTGCGAGCGCAAGGGGAGATAGCTTTACTCAAGACGCGAGCTCAAGAGGCGCATCTGG	1549
QY	275	----LeuValLeuGlyLeuLysAspLysGluProLysAlaGlnLeuSerGluLysAsnAs	293
DB	1550	AAATCAACATGAGGATCATGCGATTGTCAAGCTGAAGCAGCAGCAGCAGCAGCAGCA	1609
QY	293	pTyrLeuGlnAlaTyr-----LysAlaIleLys	302
DB	1610	GCAGCAGCAACAGCAACAGCAACAGTCTGTGACACAAACACACAGCCAAAGTCAACCAA	1669
QY	302	sSerAlaSerTyrSerGlyAlaAlaLeuAsnAlaGlyLysArgMetAlaGlyLeuProLe	322
DB	1670	AGCAGCAGCGGCAAGAGAGCAGCAGCAGCAGCAGCAGCAGCATCCAGAGTTGCAGCA	1721
QY	322	uAspMetAlaThrAspAlaMetGlyAlaValArgSerLeuValSerAlaSerSerLeuTh	342
DB	1722	---ACAGCAACAGCAGCA-----	1736
QY	342	rGlnAsnGlyLeuAlaLeuAlaGlyGlyPheAlaGlyValGlyLysLeuGlnLysMetAl	362
DB	1737	-----GC	1738
QY	362	aThrLysAsnIleThrAspProAlaThrLysAlaAlaValSerGlnLeuThrAsnLeuAl	382
DB	1739	AACAGCAACAGCAGCAACAGCAACAGCAACAGCAACAGCAACAGCAACAGCAACAGC	1798
QY	382	agLysSerAlaAlaValPheAlaGlyTyrThrThrAlaAlaLeuThrThrAspProAlaVa	402
DB	1799	AACACAGCAGCAGCAGCAGCAGCAACAGCAACAGCAACAGCAACAGCAACAGCAGCAGC	1858
QY	402	lLysLysAlaGlySerPhe-IleGlnAspThrValLysSerThrAlaSerSerThrThrG	422
DB	1859	AACTGCAGAGATCATCAGCGTGAAGATCAGGTG-----	1893
QY	422	lTyrValAlaAspGlnThrValLysLeuAlaLysThrValLysAspMetGlyGlyGluA	442
DB	1894	----ATTGAGCTCTGTACCAAGCGAGCTGGCCAAGTTCGCCAAGAGAGCTCTCCGATCTGG	1948
QY	442	lIleThrHisThrGlyAlaSer-----LeuArgAsnThrV	454
DB	1949	CCATTGCCACAGAGTACGGGAGGAGGCCCTGGCGCCGATACACTCGTCTCAAGCAGCAGC	2008
QY	454	aLysAsnLeuArgGln-----ArgProAlaArgGluAlaAspIleGluGlyGlyT	472
DB	2009	TGGACAATCTTAACGAGATTTCGCCAGAGAAGACGAAAGTACACACA-----A	2056
QY	472	hAlaAlaSerProSerGluIle	479
DB	2057	CATCCGCATCACCAACAGCAGTA	2079
RESULT 15			
AAV21187			
ID	AAV21187	standard; DNA; 53789 BP.	
XX	AAV21187;		
AC	AAV21187;		
XX			
DT	24-JUL-1998	(first entry)	
XX			
DE	Amycolatopsis mediterranei rifamycin synthesis gene cluster fragment.		
XX			

KW Amycolatopsis mediterranei; rifamycin; synthesis; gene cluster;
XX polyketide synthase; actinomycete; ansamycin; ds.
OS Amycolatopsis mediterranei.
FH Key Location/Qualifiers
FT CDS 1825..15543
FT /tag= a
FT /label= ORF_A
FT /product= "polyketide synthase"
FT CDS 15550..30759
FT /tag= b
FT /label= ORF_B
FT /product= "polyketide synthase"
FT CDS 30895..36060
FT /tag= c
FT /label= ORF_C
FT /product= "polyketide synthase"
FT CDS 36259..41325
FT /tag= d
FT /label= ORF_D
FT /product= "polyketide synthase"
FT CDS 41373..51614
FT /tag= e
FT /label= ORF_E
FT /product= "polyketide synthase"
FT CDS 51713..5293
FT /tag= f
FT /label= ORF_F
FT /product= "polyketide synthase"
PN WO9807868-A1.
XX 26-FEB-1998.
PD 18-AUG-1997; 97WO-EP04495.
PF 20-AUG-1996; 96EP-0810551.
PR (NOVS) NOVARTIS AG.
PA Engel N, Schupp T, Toupet C;
PI WPI; 1998-169172/15.
XX DR P-PSDB; AAW52845-W52850.
PT Amycolatopsis mediterranei rifamycin synthesis gene cluster - used
PT to produce rifamycin and rifamycin analogues
PS Claim 4; Page 53-102; 205pp; English.
XX The present sequence represents a Amycolatopsis mediterranei rifamycin
CC synthesis gene cluster DNA fragment from the present invention. The
CC DNA fragment comprises a DNA region involved directly or indirectly
CC in the gene cluster responsible for rifamycin synthesis, including
CC the adjacent DNA regions to the right and left which, by reason of
CC their function in connection with rifamycin biosynthesis, qualify
CC as constituents of this rifamycin gene cluster, and functional
CC fragments, derivatives or constituents of these. The Amycolatopsis
CC mediterranei rifamycin synthesis gene cluster DNA fragment can be used
CC for producing rifamycin, rifamycin analogues or precursors. It can also
CC be used for inactivating or modifying genes involved in ansamycin or
CC rifamycin biosynthesis. The DNA can be used for constructing mutant
CC actinomycetes strains from which the natural rifamycin or ansamycin
CC biosynthesis gene cluster has been partly or completely deleted. The
CC DNA fragment can be used for assembling a library of polyketide
CC synthases, which can be used for assembling a library of polyketides.
CC A hybridisation probe of the invention can be used for identifying DNA
CC fragments involved in the biosynthesis of ansamycins.

Alignment Scores:

Pred. No.: 11.7 Length: 53789
Score: 141.00 Matches: 141
Percent Similarity: 34.29% Conservative: 63
Best Local Similarity: 23.70% Mismatches: 192
Query Match: 5.85% Indels: 200
DB: 19 Gaps: 29
US-09-825-414-7 (1-486) x AAV21187 (1-53789)
QY 11 ProValThrAlaThrAspSerPheArgThrAlaSerAspAlaSerLeuAlaSerSer 30
Db 34264 CCGGTCTCACCATGAGACTCGCTG-----ATC 34290
QY 31 ValArgSerValSerSerAspGlnGlnArgGluIleAsnAlaIleAlaAspTyrLeuThr 50
Db 34291 CTGCGTGAAGTCGCCCTCGACCAAGCTGACACTGCGCGCGGCTCGCTTAACCGGTG 34350
QY 51 AspHisValPheAlaAlaHisLysLeuProProAlaAspSer----- 64
Db 34351 GACTGGACG-----CCACTGCCCACTGTGACAGTGGCGGTGCCGTGTCGG 34398
QY 65 -----AlaaspGlyGlnAlaAlaValAspValHisAsnAlaGlnIleThr 79
Db 34399 GCCGAGGTGCTGAAGCTTTCCGCGAGAGAGCCCTTGACCTGACCGCGCGGTGCTGCC 34458
QY 80 AlaLeu-----IleGluThrArg 85
Db 34459 GCCCTGCAGCGGTGCTTCCGACGCGCGGAGAGACCCGCTGTGCTGTGACCGCGG 34518
QY 86 AlaSerArgLeuHisPheGluGlyGluThrProAla-----ThrIleAlaasp--- 101
Db 34519 -----GGTGGGTGCCCGCGGAGACGGGTGTGCTGATCCG 34557
QY 102 -----ThrPheAlaLysAlaGluLysLeuAspArgLeu 112
Db 34558 GCGGGTCCGCGGTGTGGGCTGTGTCGGGGCGCGGAGAGAACCCGACCGGTTC 34617
QY 113 -----AlaThrThrThrSerGly----- 118
Db 34618 GTCTGCTCGACACGAGCGGAGGTGCCGTGAAGCGGTGCGACCGGTGAGCCG 34677
QY 119 -----AlaLeuArgAlaThrProPheAlaMetAlaSerLeuGlnTyrMetGlnPro 136
Db 34678 CAGCTCGCGCTGCGCGCACAGACGTTCTCGGTGCCCGCGCTGCACCAACCG 34737
QY 137 AlaIleAsnLysGlyAspTyrLeuProAlaProLeu-----LysPro---LeuThrPro 153
Db 34738 GCG-----GAAGCCCCGTGACGTTCCGTCGCGACGGAGCGGCTC 34776
QY 154 LeuIleSerGlyAlaLeuSerGlyAlaMetAspGlnValGlyThrLysMetMetAspArg 173
Db 34777 CTGGTCTCCGGCGCC----- 34791
QY 174 AlaThrGlyAspLeuHisTyrLeuSerAlaSerProAspArgLeuHisAspAlaMetAla 193
Db 34792 -----GGAGCGCTGGGTGCGCTGCCCGCGGACCTGTCACCGCGGACGGCGTCCGG 34845
QY 194 AlaSerValLysArgHisSerProSerLeuAlaArgGlnValLeuAspThrGlyValAla 213
Db 34846 CGGCTGCTGTCGCCAGCGCGCGCGCGCGCGCGCGCGCATGACGACCTGTCGCC 34905
QY 214 ValGlnThrTyrSerAlaArgAsnAlaValaLysArgThrValLeuAlaProAlaLeuAlaSer 233
Db 34906 GAGCTGACC-----GGCACGCGCGCGGAGTGAAGGTGCGCGCTGCGACGTCCTCCGAC 34959
QY 234 ArgProAlaValGlnGlyAlaValaAspLeuGlyValSerMetAlaGlyGlyLeuAlaAla 253
Db 34960 CGCGACCAAGTGGC-GGGCT-----GCTCAAGGAACACGCGCT 34997
QY 254 AsnAlaGly-----PheGlyAsnArgLeuLeuSerValGlnSerArgAspHisGlnArg 271
Db 34998 GACCGCGGTGTGCACACGCGCGGCGT-----GTTGCAGCGCGGTGTACCGCGCGC 35048

QY	272	Gly-----GlyAlaLeuValLeuglyLeuLysAspLys-----GluProLys	285
Db	35049	GCTGACCCGGGAGCGGCTGGCCAAAGTGTTCGCCCAAGTTCGACGC GGCCAACCACCT	35108
QY	286	Ala-GlnLeuSerGluGluAsnAspTrpLeuGluAlaTyrlLysAlaIleLysSerAlase	305
Db	35109	CGACGAGCTGAACCCGCGACCTGGAC---CTCAGACGGTTTCATCTACTGTCTCCGCCCTC	35165
QY	305	r-----TyrSerGlyAla-----Al	310
Db	35166	CTCGATCTTTCATGGGCGCGGGCACGGCGGGGTACCGGGCGGAACGCTTACTCTCGACGG	35225
QY	310	aleuAsnAlaGlyLysArgMetaGlyLeuPro--LeuAspMeta-----	325
Db	35226	CCTGATGGCCCGCCCGCGCGCGGGGCTGTCTGCTTGCCCTGGGGGCCGTG	35285
QY	325	-----	325
Db	35286	GGAGCAGCTCACCGCATGGCCGACACCATCGACGACCTCACCCCTGGCCCGGATGAGCCG	35345
QY	326	-ThrAspAlaMetGlyAlaValArgSerLeuValSerAlaSerSerLeuthrGlnAsnGl	345
Db	35346	GCGCGAAGGCCGCGGCGGCTCCGCGCTCGGCTCCGCCGACGCGCATGAGCTGTTGCA	35405
QY	345	yLeuAlaLeuAlaglyLy-----	351
Db	35406	CGCCGCGCTCGCGCGCGGACAGCGCTGCTGTCGCCGATCGAGCTCGACCTGCCGAGGT	35465
QY	352	-----PheAlaGlyValGlyLysLeuGlnLumetalThrlLysAsnIleThras	368
Db	35466	GCGGCGCCAGCGCGCGCGCGCGCGCACGCTGCCACCTG---CTGCGCGGCTGTCCG	35522
QY	368	PrroAlaThrlLysAlaAlaValSerGlnLeuThrasnLeuAlaGlySerAlaAlaValPh	388
Db	35523	CGCGGCGCGCGCGCGCGCGGACGCGGCCACCGAGAGCGGCGCTGGAACGCGGCT	35582
QY	388	ealaglyTrpThrThralaAlaLeuThrThraspproAlaValLysAlacluserPh	408
Db	35583	GGCGGGCTCACCGTGCC-----GAACAAGAAGCGCT	35615
QY	408	eileGlnAspThrValLysSerThralaSerSerThrThrglyTyrrValaAlaaspGlnTh	428
Db	35616	GCTGCTCGACCTCGTCCGCGGTCAAGTTCGCCGTCGTCTCGGGCACGCCGACACTCCGG	35675
QY	428	rValLysLeuAlaLysthrValLysaspmetGlyGlyAlaIlethrHstHrglyAl	448
Db	35676	CGTCCGCGCGGACGCGGCGTTCAAAGACGCCGGGTTTCGACTCGCTG----ACGTGGGT	35729
QY	448	aSerLeuArgAsnThrValAsnasnLeu--ArgGlnArgProAlaArggluaAlaaspIle	467
Db	35730	GGAGCTGCGCAACCGGCTGCGCGAGACGACCGGCTGAACACTGCCCGGACGCTGTCTT	35789
QY	468	-----GluGluGlyGlyThralaAlaSerProser	477
Db	35790	CGACCATCCGAACCCGCTGGCAGCTGGCGCGCGGACACCTGCG	35828

Search completed: January 31, 2003, 05:29:09
Job time : 3263.75 secs

Gencore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: January 31, 2003, 03:21:14 : Search time 1572.38 Seconds
(without alignments)
5005.784 Million cell updates/sec

Title: US-09-825-414-7
Perfect score: 2412
Sequence: 1 MHINRVQQPVPVTATDSFRT.....IEEGTAASPSSEIPRPMRS 486

Scoring table:
BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 16154066 segs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp
-Q/cgn2.1/USPTO.spool/US09825414/runat_28012003_161742_26093/app_query.fasta_1.1294
-DB=EST -QFMT=fastap -SUFFIX=p2n.rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=biosum62 -TRANS=human40.cdd -LIST=45
-DOCCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09825414@CGN_1_1_2221@runat_28012003_161742_26093 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MAP -LARGQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST:*
1: em_estba:*
2: em_esthum:*
3: em_estlin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_est1:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: gb_gss:*
18: em_gss_hum:*
19: em_gss_inv:*
20: em_gss_pln:*
21: em_gss_vrt:*
22: em_gss_fun:*
23: em_gss_mam:*
24: em_gss_mus:*
25: em_gss_other:*
26: em_gss_pro:*
27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query		DB	ID	Description
		Match	Length			
C 1	126.5	5.2	1138	17	CNS06QEO	AL410542 T3 end of
2	126	5.2	1753	11	BC018800	BC018800 Homo sapi
3	125.5	5.2	4001	11	AK014542	AK014542 Mus muscu
C 4	119	4.9	2191	17	AQ012154	AQ012154 8758PID03
C 5	115.5	4.8	883	12	BG369810	BG369810 HVSME1002
C 6	114	4.7	1792	11	AY103948	AY103948 Zea mays
7	113.5	4.7	4088	11	BC021514	BC021514 Mus muscu
8	112.5	4.7	2012	11	AY107104	AY107104 Zea mays
9	111.5	4.6	975	9	AL545051	AL545051 AL545051
C 10	111	4.6	948	13	BM044579	BM044579 603622231
11	110.5	4.6	711	13	BM588321	BM588321 170006873
C 12	110.5	4.6	888	17	CNS076CP	AL431231 T7 end of
C 13	110	4.5	860	14	BQ424977	BQ424977 AGENCOURT
C 14	109.5	4.5	670	14	BU027919	BU027919 OHG8M04.y
15	109.5	4.5	2347	11	BC020083	BC020083 Mus muscu
C 16	109	4.5	671	13	BT725871	BT725871 1031081F0
17	109	4.5	1295	12	BF973453	BF973453 602242455
18	107	4.4	3264	11	AK004694	AK004694 Mus muscu
19	106.5	4.4	1418	11	AK019388	AK019388 Mus muscu
20	106	4.4	2372	13	BM042670	BM042670 603616017
21	105.5	4.4	1038	17	CNS029PG	AL187549 Tetradon
22	105	4.4	989	14	BQ065543	BQ065543 AGENCOURT
23	104.5	4.3	1059	13	BM557363	BM557363 AGENCOURT
C 24	104.5	4.3	1108	13	BM548553	BM548553 AGENCOURT
25	104.5	4.3	1150	17	CNS079ZS	AL435950 T3 end of
C 26	104.5	4.3	1603	11	AY103757	AY103757 Zea mays
27	104.5	4.3	2802	11	BC023619	BC023619 Homo sapi
28	104.5	4.3	3441	11	BC030898	BC030898 Mus muscu
29	104	4.3	660	13	BJ476386	BJ476386
C 30	104	4.3	665	14	BQ841850	BQ841850 WHE2986.A
C 31	104	4.3	915	14	BQ952432	BQ952432 AGENCOURT
32	104	4.3	1159	17	CNS076BD	AL431075 T7 end of
C 33	104	4.3	1923	11	AY106615	AY106615 Zea mays
C 34	103.5	4.3	718	13	BT523897	BT523897 603051636
C 35	103.5	4.3	759	9	AJ497335	AJ497335
C 36	103.5	4.3	1033	14	BQ685858	BQ685858 AGENCOURT
C 37	103	4.3	774	13	BT556804	BT556804 603238734
38	103	4.3	910	14	BQ688626	BQ688626 AGENCOURT
C 39	103	4.3	1002	17	AG113524	AG113524 Pan trogl
40	103	4.3	1031	12	BG482132	BG482132 602528034
C 41	103	4.3	1979	11	AY104979	AY104979 Zea mays
C 42	103	4.3	4272	11	BC016144	BC016144 Homo sapi
43	102.5	4.2	732	14	BQ804418	BQ804418 WHE3554.D
44	102.5	4.2	746	17	AZ624776	AZ624776 IM0463K14
45	102.5	4.2	950	9	AL541558	AL541558 AL541558

ALIGNMENTS

RESULT 1
CNS06QEO/c
LOCUS
DEFINITION
CNS06QEO 1138 bp DNA linear GSS 05-JUL-2001
T3 end of clone AM0AA005C02 of library AM0AA from strain CLIB 89 of
Yarrowia lipolytica, genomic survey sequence.
ACCESSION
AL410542
VERSION
AL410542.1 GI:12179004
KEYWORDS
GSS.
SOURCE
Yarrowia lipolytica.
Yarrowia lipolytica.
ORGANISM
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Dipodascaceae; Yarrowia.
REFERENCE
1 (bases 1 to 1138)
Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,
Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,

TITLE
JOURNAL

deMontigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,
Malpertuy,A., Neuvéglise,C., Ozier-Kalogeropoulos,O., Pothier,S.,
Saurin,W., Tekala,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
Wincker,P. and Weissenbach,J.
Genomic exploration of the hemiascomycetous yeasts: 1. A set of
yeast species for molecular evolution studies
FEBS Lett. 487 (1), 3-12 (2000)

JOURNAL	FEBS Lett. 487 (1), 3-12 (2000)
MEDLINE	20584711
PUBMED	11152876

REFERENCE 2 (bases 1 to 1138)
AUTHORS Casaregola, S., Neuveglise, C., Lepingle, A., Bon, E., Feynerol, C., Artiguenave, F., Wincker, P. and Gallardin, C.

TITLE Genomic exploration of the hemiascomycetous yeasts: 17. *Yarrowia lipolytica*
JOURNAL FEBS Lett. 487 (1), 95-100 (2000)

JOURNAL	FEBS Lett. 487 (1), 95-100 (2000)
MEDLINE	20584727
PUBMED	11152892
REFERENCE	3 (bases 1 to 1138)

**JOURNAL
TITLE**

COMMENT

Submitted (07-SEP-2000) Genoscope - Centre National de Séquençage, 2 rue Gaston Crémieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail : segreff@genoscope.cns.fr - Web : www.genoscope.cns.fr) This GSS is part of a random genomic sequencing program of thirteen yeast species: *Saccharomyces bayanus* var. *uvorum*, *Saccharomyces exiguus*, *Saccharomyces servazzii*, *Zygosaccharomyces rouxii*, *Saccharomyces kluyveri*, *Kluyveromyces thermotolerans*, *Kluyveromyces lactis* var. *lactis*, *Kluyveromyces marxianus* var. *marxianus*, *Pichia angusta*, *Debaryomyces hansenii* var. *hansenii*, *Pichia sorbitophila*, *Candida tropicalis* and *Yarrowia lipolytica*. Genomic inserts of 3 to 5 kb were prepared and both extremities were sequenced. See keywords for description of this sequence and for the sequence of the other extremity of this insert.

FEATURES	Location/Qualifiers
source	1. .1138

```

/db_xref="taxon:4952"
/clone="AW0AA005C02"
/clone_lib="AW0AA"
/note="end : T3"

```

BASE COUNT	186 a	294 c	381 g	273 t	4 others
ORIGIN					

Alignment Scores:	
Pred. No.:	0.0175
Score:	126.50
Percent Similarity:	35.01%
Best Local Similarity:	24.26%
Query Match:	5.24%
DB:	17
	Gaps: 24
	Length: 1138
	Matches: 106
	Conservative: 47
	Mismatches: 169
	Indels: 115

US-09-825-414-7 (1-486) x CNS06QEO (1-1138)

OY	61	ProIaaspSeraIaaspGlyGlnAlaAlaValaspValHisAsnAlaGlnIleThrAla	80
			::: :::
Db	1097	CCAGCATCATCCGCGATCCACACCACCTCGCC-----GCAGACTCACYTCC	1053

Qy	81	LeuIleGIuThrArgAlaSerArgLeuHisPheGluGluThrProAlaThrIleAla	1000
Db	1052	CATGCTGCACGACCCAGCACCTTTCCACATGCGCGCCCATGTACCAACAACAGCAGCC	9933

OY 101 AsprthrphaalalysalaagluysleuasparglualaathrrthrrSerGlyalaleu 120
::: ||| | 11111 111111:::
Db 992 CCATCCCGCAGCATCACACCC-----CCCAACCACCGCATCCGGGTCA--- 951

Qy	121	ArgAlaThrProPheAlaMetAlaSerLeuLeuGlnTyrMet---	GlnProAlaIleAsn	139
Db	950	CATGTCGCCCCAGGGTCATCC-----CAATPTGTCTCCTCACC	CGGTCAGCA	903

QY	140	LysGlyaspT	ProLeuPro	AlaPro	LeuLys	ProLeuThr	ProLeuIle	SerGly	AlaLeu	159
							:::			
Db	902	TCAGGCTCTCATCCCGGCCAC	CACGAGGCCAC	CCAGGACAC	CCCTGGCCATC	AGGCTCA	---			846

[illegible]

```

QY 195 SerValLysArgHisSerProSerLeuAla-----ArgGlnValLeuAspThrGlyVal 212
      ||||| ||| |||::: ||| ||||| |||
DB 761 -----TCCCATTCAGGGTCACAGCTCTGTTGACGGCCATTCACCTCGCACCGSAGTC 711

```

213 AlavAlGlnThrTyrSerAlaArgGlnAlaValAlaArgThrValLeuAlaProAlaLeuAla 232

			:::	
Db	710	ACC-----ACCCGGGATGCAGATGATCC-----CGC	684	
QY	233	SerArgProIaIaValGlnGlyAlaValaIaAspLeuGlyValSerMetIaGlyGlyLeuAla	252	

Db 683 ATACGACAGCCCAAGACGGGATTTCTACAGCAGTCCACAGCCGGCTGGAGGCCCTGGT 624

Db 623 GCA---GCTGCA-----AGAGTTGCGAGACAGGTCGTGATGATGAATACAC 579

QY 270 GlnArgGlyGlyAlaIeuValIeuGlyLeuLysAspLysGluProLysAlaGlnIeuSer 289
::: |||::: ::: |||

290 G L G L I u a s n a s p T r P l e u g l u a l a t y r l y s a l a i l e l y s s e r a l a s e r t y r s e r g l y a l a 309

Db 548 -----TACGCA GCCGTTAAGCAAAAGCGCTAAGACGTCCTG 513

Db 512 CAAAGGAGCGCATTTGTCGGTGTGGAGTACGAGAGGGTGTGCGCTGGCC 453

QY 330 GYLAVAlArGerLeuValSerAlaSerSerLeuThrgInAnglyLeuAlaLeuAla 349
||||| ||| ||||:|||| :|| :|||
||||| ||| ||||:|||| :|| :|||

350 G1G1YPHeaIaGlyValG1YLysLeuG1uMetalaThrIlyAsn1IeThraSPPro 369

Db 393 ----- 393

Db 392 TCCACATCGTAGCACCAGCGTACGTTCTCTGCTGCAGAGCGGCTGCCGCCGCTGCCGCC 333

QY 390 G L Y T R P T h r t h r a l a l e u t h r t h r a s p r o a l a v a l y s l y s a l a g l u s e r p h e i l e 409
|||||:||||:||||| ||| ||||| :|||

410	GlnsptThr-----ValLysSerThrAlaSerSerThrThrgLgTYrValAlaaspGln	427
-----	---	-----

Db 278 TCTAGACGTACGCTGCTGCTTCTGTTCGACGCGGAACCTCTCTAAGGAGGCTTCCCT 219

	420		:		
	421		:		
	422		:		
	423		:		
	424		:		
	425		:		
	426		:		
	427		:		
	428		:		
	429		:		
	430		:		
	431		:		
	432		:		
	433		:		
	434		:		
	435		:		
	436		:		
	437		:		
	438		:		
	439		:		
	440		:		
	441		:		
	442		:		
	443		:		
	444		:		
	445		:		
	446		:		
	447		:		
	448		:		
	449		:		
	450		:		
	451		:		
	452		:		
	453		:		
	454		:		
	455		:		
	456		:		
	457		:		
	458		:		
	459		:		
	460		:		
	461		:		
	462		:		
	463		:		
	464		:		
	465		:		
	466		:		
	467		:		
	468		:		
	469		:		
	470		:		
	471		:		
	472		:		
	473		:		
	474		:		
	475		:		
	476		:		
	477		:		
	478		:		
	479		:		
	480		:		
	481		:		
	482		:		
	483		:		
	484		:		
	485		:		
	486		:		
	487		:		
	488		:		
	489		:		
	490		:		
	491		:		
	492		:		
	493		:		
	494		:		
	495		:		
	496		:		
	497		:		
	498		:		
	499		:		
	500		:		
	501		:		
	502		:		
	503		:		
	504		:		
	505		:		
	506		:		
	507		:		
	508		:		
	509		:		
	510		:		
	511		:		
	512		:		
	513		:		
	514		:		
	515		:		
	516		:		
	517		:		
	518		:		
	519		:		
	520		:		
	521		:		
	522		:		
	523		:		
	524		:		</

QY 448 AlaserleuArgAsnThrValAsnAsnleuArgGlnArgProAlaArgGluAlaaspIle 467
::: ||| ||||| :::

0Y 468 GJUGJGVGIvThrAlaAAserProSergIuIePropheArqPromet 484
DD 1/5 -----CAAGGAGGAAAGACAAAGGAGAAIAGGGAAAGAAAATGC 191

Db 140 GCGGCTAAGCAAGCTGCGGCTSTGCCAAGCAGCTTG--ATCAGGCTATG 93

RESULT 2	
BC018800	
TCCTTC	
BC018800	
1752 hr	
mpna	
1 linear	
MMC 07-DEC	

DEFINITION Homo sapiens, clone IMAGE:3605889, mRNA.


```

ACCESSION      BC018800
VERSION        BC018800.1  GI:17402947
KEYWORDS
SOURCE         Homo sapiens.
ORGANISM       Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 1753)
AUTHORS        Strausberg, R.
TITLE          Direct Submission
JOURNAL        Submitted (07-DEC-2001) National Institutes of Health, Mammalian
               Gene Collection (MGC), Cancer Genomics Office, National Cancer
               Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
               USA
REMARK         NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT        Contact: MGC help desk
               Email: cga@bbs-r@mail.nih.gov
               Tissue Procurement: ATCC
               cDNA Library Preparation: Rubin Laboratory
               cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
               DNA Sequencing by: Genome Sequence Centre,
               BC Cancer Agency, Vancouver, BC, Canada
               info@bcgsc.bc.ca
               Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,
               Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,
               Leticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo
               Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven
               Ness, Pawan Pandoh, Anna-Lissa Prabhu, Parvaneh Saeedi, Jacqueline
               Schein, Duane Smalls, Michael Smith, Lorraine Spence, Jeff Stott,
               Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy,
               George Yang, Scott Zuyderduyn, Marco Marra.
FEATURES       location/Qualifiers
               source
               1..1753
               /organism="Homo sapiens"
               /db_xref="taxon:9606"
               /clone="IMAGE:3605889"
               /tissue_type="uterus, endometrium adenocarcinoma"
               /clone_lib="NIH-MGC_44"
               /lab_host="DH10B-R"
               /note="Vector: pOTB7"
BASE COUNT    334 a 458 c 558 g 403 t
ORIGIN
Alignment Scores:
Pred. NO.:    0.0379          Length:      1753
Score:        126.00         Matches:     91
Percent Similarity: 38.07%    Conservative: 59
Best Local Similarity: 23.10% Mismatches:   139
Query Match:  5.22%         Indels:     106
DB:           11           Gaps:         16
US-09-825-414-7 (1-486) x BC018800 (1-1753)
QY   53 ValPheAlaAlaHIsLysLeuProProAlaAspSerAlaAspGlyGlnAla---AlaVal 71
    ::: |||||  |||  ::: |||::|||  ||:::
Db   219 ATACGGCGCGGACGCCGAAACCTCGAGGCTCGCAGCGCGGGGCCAAGGCCCGGCTTG 278
    ::::: |||::|||::: |||  :::::
QY   72 AspValHisAsnAlaGlnIleThrAlaLeu-----IleGluThrArg 85
    ::::: |||::|||::: |||  :::::
Db   279 GAGATGGAGAATTCTCAGTTGTGTAGCTGTTCAATCGCGCGCTCAATGTGCAGCAGT 338
    ::::: |||::|||::: |||  :::::
QY   86 AlaSerArgLeuHisPheGluGlyGluThrProAlaThrIleAlaasp----- 101
    |||  |||  |||  |||::: |||
Db   339 GAGTCGGGCGCTGCGCGCCACTTTGAGGCTTTGGGACTCTGACGCGACTGCGTGGTGTG 398
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY   102 -----ThrPheAlaLysAlaGlu 107
    |||::: |||

```

Db	Accession	Protein	Start	End	Length	Score	E-value	Model
Db	399	GTGAATCCCCAGACCAAGCGCTCCCGTTGCTTTGGCTTCGTACCTACTCCAAATGTGAG	127	146	20	1.0	1.0	1.0
QY	108	LysLeuAspArgLeuAlaThrThrThrSerGlyAlaLeuArgAlaThrProPheAlaMet	127	146	20	1.0	1.0	1.0
Db	459	GAGCGCGAC-----GCCGCCATGGCGCCCTCGCCCATGGCCGTG	497	516	20	1.0	1.0	1.0
QY	128	AlaSerLeuLeuGlnTyrMetGlnProAlaIleAsnLysGlyAspTrpLeu--ProAla	146	165	20	1.0	1.0	1.0
Db	498	GACGGCAACACTGTGGAGCTGAAGCGGGCGGTGTCGCCGAGAGATTCGGCGGGCCCGGT	557	576	20	1.0	1.0	1.0
QY	147	ProLeuLysProLeuThrProLeuIleSerGlyAlaLeuSerGlyAlaMetAspGlnVal	166	185	20	1.0	1.0	1.0
Db	558	GCCCAACCGCAAGGTTAAGAACTCTTGTGCGAGGCGCTTAAGAAG-----	602	621	20	1.0	1.0	1.0
QY	167	GlyThrLysMetMetAspArgAlaThrGlyAspLeu-----HisTyr-LeuSerAlaSe	184	203	20	1.0	1.0	1.0
Db	603	-----GACGTGGCTGAGGGCGGACCTGATGAGACACTTCTCGCAGTTTGGC	647	666	20	1.0	1.0	1.0
QY	184	rProAspArgLeuHisAspAlaMetAlaAlaSerValLysArgHisSerProSerLeuAl	204	223	20	1.0	1.0	1.0
Db	648	ACCGTGGAAAGCGCCGAGATTATTGCCGACAAG-----AGTCCGGC	689	708	20	1.0	1.0	1.0
QY	204	ArgGlnValLeuAspThrGlyValAlaValGlnThrTyrSerAlaArgAsnAlaValArg	224	243	20	1.0	1.0	1.0
Db	690	AAGAAGCGTG--GATTCGGCTTCGTGATTTCCAGATACAGACGGCGGAG-----	738	757	20	1.0	1.0	1.0
QY	224	GThrValLeuAlaProAlaLeuAlaSerArgProAlaValGlnGlyAlaVal--AspLeuG	244	263	20	1.0	1.0	1.0
Db	739	-----ACAAGCGCGGTGTCAGTTCCATCCGATTCCAG	773	792	20	1.0	1.0	1.0
QY	244	IlyValSerMetAlaGlyGlyLeuAlaAlaAsnAlaGly-----PheGly-	258	277	20	1.0	1.0	1.0
Db	774	GGCCATCGCGGTGAGGTGAAGAAAGCAGTCCCAAGAGAGATATCTACTCCGTTGGGGT	833	852	20	1.0	1.0	1.0
QY	259	--AsnArgLeuLeuSerValGlnSerArgAspHisGlnArgGlyGlyAlaLeuValLeuG	278	297	20	1.0	1.0	1.0
Db	834	GGAGCGGCTCCGATCCCTCCCGGGCGCGGCGGAGCGGCC--GGGGCGCGCGGTGTGC	892	911	20	1.0	1.0	1.0
QY	278	LylLeuLysAspLysGlnProLysAlaGlnLeuSerGlnGluAsnAspTrpLeuGlnAlaTr	298	317	20	1.0	1.0	1.0
Db	893	AGACCAGAACGGCTTTCCAAGGGCGG-----	921	940	20	1.0	1.0	1.0
QY	298	YrlYsAlaIleLysSerAlaSerTyrSerGlyAlaAlaLeuAsnAlaGlyLysArgMetAla	318	337	20	1.0	1.0	1.0
Db	922	--CGCGCGGTACACACAGCTACGGTGTACGGCGCGCGCGGAGCGGCG-----	969	988	20	1.0	1.0	1.0
QY	318	IagLylLeuProLeuAspMetAlaThrAspAlaMetGlyAlaValArgSerLeuValSerAla	338	357	20	1.0	1.0	1.0
Db	970	-----GCTACATGCTCTACGAGGCGGGCGGGCGGCTTCCTACGGTGGAG	1018	1037	20	1.0	1.0	1.0
QY	338	IaSerSerLeuThrGlnAsnGlyLeuAlaLeuAlaGlyGlyPheAlaGlyValGlyLysL	358	377	20	1.0	1.0	1.0
Db	1019	CGACTACGGTAAACGGCTTCGGCGGCTTCG-----	1047	1066	20	1.0	1.0	1.0
QY	358	eugGlnGluMetAlaThrLysAsnIleThrAspProAla-----ThrLysAlaAlaVal	375	394	20	1.0	1.0	1.0
Db	1048	-----GCAGCTACAGCCACGATCAGTCTCTATGGGCCCATGAAGAGCGGCGCGG	1099	1118	20	1.0	1.0	1.0
QY	375	AlSerGlnLeuThrAsnLeuAlaGlySerAlaAlaValPheAlaGlyTrpThrThrAlaA	395	414	20	1.0	1.0	1.0
Db	1100	CGGCGCGGTGAGGAGCAGTACCTGGGGCGGTGCGCAGTATAGTGGACCTTACAGAGCGG	1159	1178	20	1.0	1.0	1.0
QY	395	IaLeuThrThrAspProAlaValLysLysAlaGluSer	407	426	20	1.0	1.0	1.0
Db	1160	CTATGGCGGTGGGGGTGCTATGGAGGCAAGCTCTCTCT	1197	1216	20	1.0	1.0	1.0

VERSION	AK014542.1	GI:12852462
KEYWORDS	HTC, CAP trapper.	
SOURCE	Mus musculus (strain:C57BL/6J) 0 day neonate skin cDNA to mRNA, clone_lib:RIKEN full-length enriched mouse cDNA library clone:4631434016.	
ORGANISM	Mus musculus Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	
REFERENCE	1	
AUTHORS	Carninci,P. and Hayashizaki,Y.	
TITLE	High-efficiency full-length cDNA cloning	
JOURNAL	Meth. Enzymol. 303, 19-44 (1999)	
MEDLINE	99279253	
PUBMED	10349636	
REFERENCE	2	
AUTHORS	Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.	
TITLE	Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes	
JOURNAL	Genome Res. 10 (10), 1617-1630 (2000)	
MEDLINE	20499374	
PUBMED	11042159	
REFERENCE	3	
AUTHORS	Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P., Konno,H., Akiyama,J., Nishi,K., Kitsuunai,T., Tashiro,H., Itoh,M., Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A., Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.	
TITLE	RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer	
JOURNAL	Genome Res. 10 (11), 1757-1771 (2000)	
MEDLINE	20530913	
PUBMED	11076861	
REFERENCE	4	
AUTHORS	Kawai,J., Shinagawa,A., Shibata,K., Yoshino,M., Itoh,M., Ishii,Y., Arakawa,T., Hara,A., Fukunishi,Y., Konno,H., Adachi,J., Fukuda,S., Aizawa,K., Izawa,M., Nishi,K., Kiyosawa,H., Kondo,S., Yamanaka,I., Saito,T., Okazaki,Y., Gojobori,T., Bono,H., Kasukawa,T., Saito,R., Kadota,K., Matsuda,H., Ashburner,M., Batalov,S., Casavant,T., Fleischmann,W., Gaasterland,T., Gissi,C., King,B., Kochiwa,H., Kuehl,P., Lewis,S., Matsuo,Y., Nikaido,I., Pesole,G., Quackenbush,J., Schriml,L.M., Staubli,F., Suzuki,R., Tomita,M., Wagner,L., Washio,T., Sakai,K., Okido,T., Furuno,M., Aono,H., Baldarelli,R., Barsh,G., Blake,J., Boffelli,D., Bojunga,N., Carninci,P., de Bonaldo,M.F., Brownstein,M.J., Bult,C., Fletcher,C., Fujita,M., Gariboldi,M., Gustincich,S., Hill,D., Hofmann,M., Hume,D.A., Kamiya,M., Lee,N.H., Lyons,P., Marchionni,L., Mashima,J., Mazzarelli,J., Mombaerts,P., Nordone,P., Ring,B., Ringwald,M., Rodriguez,I., Sakamoto,N., Sasaki,H., Sato,K., Schonbach,C., Seya,T., Shibata,Y., Storch,K.F., Suzuki,H., Toyo-oka,K., Wang,K.H., Weitz,C., Wiltaker,C., Wilming,L., Wynshaw-Boris,A., Yoshida,K., Hasegawa,Y., Kawaji,H., Kohsuki,S. and Hayashizaki,Y.	
TITLE	Functional annotation of a full-length mouse cDNA collection	
JOURNAL	Nature 409 (6821), 685-690 (2001)	
MEDLINE	21085660	
PUBMED	11217851	
REFERENCE	5 (bases 1 to 4001)	
AUTHORS	Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A., Arakawa,T., Baldarelli,R., Bono,H., Brownstein,M., Bult,C., Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M., Hanagaki,T., Hara,A., Hayatsu,N., Hill,D., Hiramoto,K., Hiraoka,T., Hori,F., Hume,D., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kasukawa,T., Kato,H., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Koyama,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C., Quackenbush,J., Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D., Schriml,L., Shibata,K., Shibata,Y., Shinagawa,A., Shitaki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yamanaka,I., Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and	

TITLE	Hayashizaki,Y.
JOURNAL	Direct Submission Submitted (10-JUL-2000) Yoshinhide Hayashizaki, The Institute of Physical and Chemical Research (RIKEN), Laboratory for Genome Exploration Research Group, RIKEN Gemomic Sciences Center (GSC), RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan (E-mail:genome-res@sc.riken.go.jp, URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222, Fax:81-45-503-9216)
COMMENT	please visit our web site (http://genome.gsc.riken.go.jp/) for further details. cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in RIKEN. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues. First strand cDNA was primed with a primer [5', GAGAGAGAGAGATCCAGAGCGTCCTTTTCTTTTCTTTT 3'], cDNA was prepared by using trehalose thermo-activated reverse transcriptase and subsequently enriched for full-length by cap-trapper. cDNA went through one round of normalization to Rot = 10.0 and subtraction to Rot = 100.0. Second strand cDNA was prepared with the primer adapter of sequence [5', GAGAGAGAGATTCGAGTTAATTAAATTATCCCCCCCCCC 3']. cDNA was cleaved with BamHI and XhoI. cDNA of size comprised longer than 7 kb was selected before cloning. Vector: a modified pBluescript KS(+) after bulk excision from Lambda FLC I. Cloning sites, 5' end: SalI; 3' end: BamHI. Host: DH10B.
FEATURES	Location/Qualifiers 1..4001 /organism="Mus musculus" /strain="C57BL/6J" /db_xref="FANTOM_DB:4631434016" /db_xref="MGD:MGI:1897732" /db_xref="taxon:10090" /clone="4631434016" /tissue_type="skin" /clone_lib="RIKEN full-length enriched mouse cDNA library" /dev_stage="0 day neonate" 1..4001 /gene="Hiplr" 114..3320 /gene="Hiplr" /note="data source:MGD, source key:MGI:1352504, evidence:ISS huntingtin interacting protein 1 related putative" /codon_start=1 /protein_id="BAB29420.1" /db_xref="GI:12852463" /db_xref="MGD:MGI:1352504" /translation="MNSIKNVPAVLYSRTPGHSLFAERQFDKTOAISISKAINSOEA PVKEKHARRIILGTHHEKGFTWSYAIGLPLSSSSISMKFCHVLAKVLRDGHPNVL HDYGRKSNIREIGDLMGHLRDYGHVNIYTKLLTFKISPHLKHQFPAGLEVDEV LEKAGTDVNNIIFOLTVEMFDYMDCELKLSSEVFRQLNPAIVASOMSSGOCRLAPLIO VIQDCSHLYHYTVKLMFKLHSCLPADTLQGHDRFHEQFSLKNFFRRASDMLYFKRL IOIPRLPEGPENFLRASALAEHIKPVVVIPEAPEEEPENLIEISSAPPAGEPVVVA DLFDQTFGPNGSMKDDRDLDIENLKREVEVTLRAELKTEKMEAOYRISOLKGVNGLE AELEEQKOKOKALVDNEQLRHELAQLKALQLEGARNQGLREEAERKASATEARYSKL KEKHSDELINTHAELLRKNNADPAKQLTVTQOSQEEVARVKQOLAFOMEQAKRESEMKE EQSDQLEKLRRLARAGELARAOEALSRTQSGSELSRLDTLNAEKEALSQVRRQ EAELLAQSLVREKEALSQEOSSQEKELRGQLAERQSQEGLRQKLDEQLAVL RSAALAEAILQDAVSKLDDPLHLRCTSSPDYLVSRQAALDSVSGLEQHTQYLAAS EDASALVAALTRFSLAADTVNGAATSHLAPTDPADRLMDTCREGARALELVGOLQ DQTVLPRAQPSLMRAPLQGIQGDQDKPKSLDVROEELGAMVDKEMAATSAIEDAV RRIEDMMSQARHSSGVKLEVNERILNSCTDLMKAIRLVMTSTSLQKEIVESGRGA TQOEYPAKNSRWTEGLISASKAVWGATQLVESADKVYHMGKYELIVCSHEIAAST AQLVAASKVKANKNSPHLSRLQSCSTRVNERAANVASTKSGQEQIEDRDTMDFSGLS LILKRCOMETOVRLVLELEKTEIAERVRLGELRKOHYVLAGMGGTPSEEPSPRPAP RSGATKPPPLAQKPSIAPRTDNLDDKKDGVYPAQLVNY"
BASE COUNT	925 a 1120 c 1260 g 695 t 1 others
ORIGIN	
Alignment Scores:	

Pred. No.:	0.149	Length:	4001
Score:	125.50	Matches:	118
Percent Similarity:	33.81%	Conservative:	71
Best Local Similarity:	21.11%	Mismatches:	208
Query Match:	5.20%	Indels:	163
DB:	11	Gaps:	21
US-09-825-414-7 (1-486) x AK014542 (1-4001)			
QY	19	ArgThrAlaSerAspAlaSerLeuAlaSerSerSerValArgSerValSerSerAspGln	38
DB	1722	CGCACAGAACAGAGTGGGTCTCAGAGCTGAGCTCAGCGGCTGCACACACTGAAATGCCGAGAAG	1781
QY	39	GlnArgGluIleAsnAlaIleAlaAspTyrLeuThrAspHisValPheAlaAlaHisLys	58
DB	1782	GAA-----GCTCTGAGTGGAGTCGTCGGCAGCGT-----	1811
QY	59	LeuProProAlaAspSerAlaAspGlyGlnAlaAlaValAspValHisAsnAlaGlnIle	78
DB	1812	-----GAGGCAGAGCTGCTGGCCCTCAGAGCCCTGTGCGGGAGAGAAGAG	1856
QY	79	ThrAlaLeuIleGluThrArgAlaSerArgLeuHisPheGluGlyGluThrProAlaThr	98
DB	1857	GAGGCACTTACGCCAAGACAGCAGCGGAGCTCCAGAGAGAGGCGCAGCTACGGGGCAG	1916
QY	99	IleAlaAspThrPheAlaLysAlaGluLysLeu-----AspArgLeu	112
DB	1917	CTGGCAGAAAGAGTCTCAGAGCAGGCGCTTCGGCAGAGCTGCGATGAGCAGTTG	1976
QY	113	AlaThrThrThrSerGlyAlaLeuArgAlaThrProPheAlaMetAlaSerLeuGln	132
DB	1977	GCGGTGTGCGAAGTGCAGCCGCCGAGGCAAGGCCATC-----	2015
QY	133	TyrMetGlnProAlaIleAsnLysGlyAspTyrLeuProAla---ProLeuLysProLeu	151
DB	2016	---CTACAGGATGCAGTGAGCAAGCTGGACGA-CCCCCTGCACCTCCGCTGCACCAAGCTC	2071
QY	152	ThrProLeuIleSerGlyAlaLeuSerGly-AlaMetAspGlnValGlyThrLysMetLe	171
DB	2072	CCCAGACTACTGTGAGCCGGGCTCAGGCAAGCCCTGGACAGCGGTGAGC-----	2120
QY	171	tAspArgAlaThrGlyAspLeuHisTyrLeuSerAlaSerProAspArgLeuHisAspAl	191
DB	2121	-GGCTGGAGCAGGGCCACACCCAGTACTGTGCTTCTCTGAAGATGCT-----TCTGC	2173
QY	191	aMetAlaAlaSerValLysArgHisSerProSerLeuAlaArgGlnValLeuAspThrG1	211
DB	2174	CCTGGTGCGAGCGCTGACCCGCTTCTCC--CATTTGGCTGCGACACACCAATTGTCAATGG	2230
QY	211	yValAlaValGlnThrTyrSerAlaArgAsnAlaVal-----	223
DB	2231	TGCCGCCACCTCCACCTGGCCCCCACCAGCCCGCCGACCGCTGATGACACATGCAG	2290
QY	224	-----ArgThrValLeuAl	228
DB	2291	GGAGTGTGAGCCCGGCTCTGAGCTGGTGGGACAGCTGCAAGACCAGACAGTGTACC	2350
QY	228	aProAlaLeuAlaSer-----ArgProAlaValGlnGlyAlaValAspLeuGlyValSe	246
DB	2351	GAGGGCTCAGCCCAAGCTGATGCGGGCCCCCTGCAGGGCATTTCTGCAGTTGGGCCAGGA	2410
QY	246	rMetAlaGlyGlyLeuAlaAlaAsnAlaGlyPheGlyAsnArgLeuLeuSerValGlnSe	266
DB	2411	CTTG-----AAGCCTAAGACCTTGATGT	2434
QY	266	rArgAspHisGlnArgGlyGlyAlaLeuValLeuGlyLeuLysAspLysGluProLysAl	286
DB	2435	ACGGCAGAGAGAGTAGGG-----GCCATGTGTGACAAAGAGATGGCGGC	2479
QY	286	acLLeuSerGluGluAsnAspTyrPLeuGluAlaTyrLysAlaIleLysSerAlaSer--	305
DB	2480	CACCTCGGACGCCATTGAGGAGCGTGTGCGGAGGATCGAGACATGATGAGCCAGGCCCG	2539

QY	306	-----TyrSerGlyAlaAlaLeuAsnAlaGlyLysArgMetAlaGlyLeuProLeuAs	323
DB	2540	CCACGAGAGCTCGGGCGTGAACACTGAGGTGAATGAGAGATC-----CTCAA	2587
QY	323	pMetAlaThrAspAlaMetGlyAlaValArgSerLeuValSerAlaSerSerLeuThrG1	343
DB	2588	CTCCTGCACAGACTGATGAAGCCTATCCGCTCCTGTGATGACCTCCACCAGCCTGCA	2647
QY	343	nAsnGlyLeuAlaLeuAlaGlyGlyPheAlaGlyValGlyLysLeuGlnGluMetAlaTh	363
DB	2648	GAAGGAATTGTGGAGAGCGGC-----AGGGGGCAGCAACGACGAGCAATTTATGC	2701
QY	363	rLysAsnIleThrAspProAlaThrLysAlaAlaValSerGlnLeuThrAsnLeuAlaGl	383
DB	2702	CAAGAA-----TCACGGTGGACTGAAGCCCTCAT	2731
QY	383	ySerAlaAlaValPheAlaGlyTyrPThrThrAlaAlaLeuThrThrAspProAlaValLy	403
DB	2732	CTCAGCCTTAAGCAGTGGGCTGGGAGCCACACAGCTG-----GTGA	2776
QY	403	sLysAlaGluSerPheIleGlnAspThrValLys-----	414
DB	2777	GTCAGCTGACAAAGTGTGCTTCACATGGGCAAAATACGAGAACTATCGTCTGCCA	2836
QY	415	-SerThrAlaSerSerThrThrGlyTyrValAlaAspGlnThrValLys-----	430
DB	2837	TGAGATTGCGGCGCACAGCGGCCAGCTGTGTGCGCAGCTCGAAAGGTGAAGCCAACAAGA	2896
QY	431	-----LeuAlaLysThrValLysAspMetGlyGlyG1	441
DB	2897	CAGTCCCACCTTGAGCCCGCTGCAGGAATGTTCGCCACTGTCAACGAGAGGGCTGCCAA	2956
QY	441	uAlaIleThrHisThr-----	446
DB	2957	CGTCGTGGCCTCCACCAATCTGGCCAGAGCAGATTGAGGACAGACACCATGATTT	3016
QY	447	---GlyAlaSerLeuArgAsnThrValAsnAsnLeuArgGlnArgProAlaArgGluAl	465
DB	3017	CTCTGGCCTGTCCCTCATCAAGTTGAAGAAAGCAGAGATGAGACACAGCTGCGAGTCTT	3076
QY	465	aAspIleGluGlu-----	469
DB	3077	GGAGCTGAGAGAAGACACTAGAGCAGAGCGTGTCCGGCTCGGGAGCTTCGGAACAAGCA	3136
QY	470	-----GlyGlyThrAlaAlaSerProSerGluIleProPheArgPro	483
DB	3137	CTATGTACTGCGCTGGGGGATGGGAACACCTAGCGAAGAAGAACCCAGCAGACCC	3191
RESULT 4			
AQ012154/c			
LOCUS			
8758P1D037011098 Cosmid library of chromosome II Rhodobacter			
DEFINITION			
sphaeroides genomic clone 8758P1D037011098, DNA sequence.			
ACCESSION			
AQ012154			
VERSION			
AQ012154.1 GI:3177109			
KEYWORDS			
GSS.			
SOURCE			
Rhodobacter sphaeroides.			
ORGANISM			
Rhodobacter sphaeroides; Alpha proteobacteria; Rhodobacteriales;			
Bacteria; Proteobacteria; Rhodobacter.			
REFERENCE			
1 (bases 1 to 2191)			
Choudhary, M., Mackenzie, C., Mouncey, N., Weinstein, G.M. and Kaplan			
AUTHORS			
TITLE			
RSGDB, the Rhodobacter sphaeroides Genome Database			
JOURNAL			
Unpublished (1998)			
COMMENT			
Contact: Choudhary, M.			
Department of Microbiology and Molecular Genetics			
University of Texas Medical School			
6431 Fannin Street, Houston, TX 77030, USA			
Tel: 713 500 5437			
Fax: 713 500 5499			
Email: madhuetmng.med.utn.tmc.edu			
Seq primer: pbluescript T7			

```

Class: shotgun.
Location/Qualifiers
FEATURES
  source
    /organism="Rhodobacter sphaeroides"
    /strain="2.4.1T"
    /db_xref="taxon:1063"
    /clone="8758PID037011098"
    /clone_lib="Cosmid library of chromosome 11"
    /lab_host="E. coli S17-1"
    /note="Vector: pLA2917"

BASE COUNT      341 a      749 c      718 g      382 t      1 others
ORIGIN

Alignment Scores:
Pred. No.:      0.284      Length:      2191
Score:          119.00      Matches:      94
Percent Similarity: 37.99%      Conservative: 42
Best Local Similarity: 26.26%      Mismatches:  144
Query Match:    4.93%      Indels:      79
DB:             17      Gaps:        14

US-09-825-414-7 (1-486) x AQ012154 (1-2191)

QY  143 TTPLEUPROALAPROLEUYSPROLEUTHRPROLEULESERGLYALALeuSerGlyAla 162
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  1672 TGGCGGCGCGGCGCGCG-AAAACCTGCCGACACGCTCTGCATCGCGGTGATCTCCGCGG 1614

QY  163 MetAspGlnValGlyThrLysMetMetAspArgAlaThrGlyAspLeuHisTyrLeuSer 182
    |||:::||| |||:::||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  1613 ATGGAGCAGACCCCTGCCCGAGCTTTCGGCGCGGGTTTGGCGGAGGCTGGCGTGTGCCCG 1554

QY  183 AlaserProAspArgLeuHisAspAlaMetAlaAlaserValLysArg----- 198
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  1553 GCACGACGAGATCCGCTTAGTCACCAGCGGTGCCCTCAATCACCGCGAGACGATCGACTG 1494

QY  199 ---HisSerProSerLeuAlaArgGlnValLeuAspThrGlyValAlaValGlnThrTyr 217
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  1493 GCCTATCGTCTCT----- 1470

QY  218 SerAlaArgAsnAlaValArgThrValLeuAlaProAlaLeuAlaSerArgPro----- 235
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  1469 -----ACGAGTGGTGTCTTCAGCTGCGCACCGTCTGTCGAGCGCGTGAACC 1425

QY  236 AlaValGlnGlyAlaValAspLeuGlyValSerMetAlaGlyGlyLeuAlaAlaAsnAla 255
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  1424 GCCTCGTCTGGCGCCCTCACG----- 1386

QY  256 GlyPheGlyAsnArgLeuLeuSerValGlnSerArgAspHis-----GlnArg 271
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  1385 GCGGCCACGTCGAGCAGATGTCGTAATGCCATGGAGCAGATTCGCGCCACGCGCGCAG 1326

QY  272 GlyGlyAlaLeuValLeuGly-----LeuLysAspLysGluProLysAlaGln 287
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  1325 GCCGCTTCGCGCGGTACCGGCGCATCCAGAGCAGCCTCGAGCAGTCGGTCGAGCGCGCC 1266

QY  288 LeuSerGluGluAsnAspTyrPheGluAlaTyrLysAlaIleLysSerAlaSerTyrSer 307
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  1265 GCCTGCGAAAGCGTGGCGCGCGCGCGCGCGCAGGATGCGCGCCTCGGAGAAGTCGTT 1206

QY  308 GlyAlaAlaLeuAsn-AlaGlyLysArgMetAlaGlyLeuProLeuAspMet----- 324
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  1205 GGAGAGGCGCTCGCTTCATGAAGAAAGATCGCGGAACATATTCTCGTCATGCAGGAAT 1146

QY  325 -----AlaThrAspAlaMetGlyAlaValArgSerLeuValSerAlaSerSerLeuTh 342
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  1145 CGCGCGGCGACAGACGATCTTCTC---GCCCTGAATGCGGCGGTGGAGCGCGCGCGGG 1089

QY  342 rGlnAsnGlyLeuAlaLeuAlaGlyGlyPheAlaGlyValGlyLysLeuGlnGluMetAl 362
    ::| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  1088 CGAGCGGGGCGCTGGCGCTTTCAGTCGTGGCCAGCGAGGTCAGAAAGCTCGCCGAGCCGAG 1029

QY  362 aThrLysAsnIleThrAspProAlaThrLysAlaAlaValSerGlnLeuThrAsnLeuAl 382
    : ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```

Db  1028 C-----CAGGACGGCGCCCAAGGA-----GTCTCGACCCCTCTCCACTCGAC 987

QY  382 aglySerAlaAlaValPheAlaGlyTyrThrThrAlaAlaLeuThrThrAspProAlaVala 402
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  986 GTCCGAGCGCGCGCGTGGCGCGCGCGGATATCATTCGGCGCGCTCATCCCGGAT----- 935

QY  402 lLysLysAlaGluSerPheIleGlnAspThrValLysSerThrAlaSerSerThrThrGl 422
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  934 -----ATCGAACTCACCTTCACAGCTCGCGACC-- 908

QY  422 yTyrValAlaAspGlnThrValLysLeuAlaLysThrValLysAspMetGlyGluAl 442
    ::| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  907 -----GAGATTGCATCCGTGGCACGGGATCTGAGCGCGGAGAT 870

QY  442 alleThrHisThrGlyAla-SerLeuArgAsnThrValAsnAsnLeuArgGlnArgProA 462
    ::| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  869 CCGCTCGGTGACCGGTTGATCGACACACGCTCGACACGCTCAGCAAGCAAGATTCCTCCGT 810

QY  462 lAArgGluAlaAspIleGluGluGlyGlyThrAlaAlaSerProSerGlu 478
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db  809 ATCGG-----ACGAACCTCTGGAACGGCTCTGCCGAGCTCTCAG 772

RESULT 5
BG369810/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES
  source
    /organism="Hordeum vulgare"
    /cultivar="Morex"
    /db_xref="taxon:4513"
    /clone="HVSMEI0026C02f"
    /clone_lib="Hordeum vulgare 20 DAP spike EST library
    HVCDNA0010 (20 DAP)"
    /tissue_type="20 DAP spike"
    /lab_host="SOLR"
    /note="Vector: lambdaZAP; site_1: EcoRI; site_2: XhoI;
    Plants were grown in the greenhouse at the University of
    California, Riverside (Fenton, SJ Close, TJ Close). Whole
    spikes with awns trimmed were collected at 20 DAP (Fenton
    ). Total RNA was prepared, poly(A) RNA was purified, one
    primary unamplified cDNA library was made, and 1 million
    pfu were in vivo excised to give pBluescript SK(-) cDNA
    phagemids in the TJ Close lab at the University of
    California, Riverside (Choi). Phagemids were plated and
    picked at the Clemson University Genomics Institute (CUGI)
```


(Begum, Palmer, Frisch, Atkins and Wing). Plasmid DNA preparations, DNA sequencing and sequence analysis were performed at CUGI (Wing, Yu, Frisch, Henry, Simmons, Oates, Rambo, Main). The sequence has been trimmed to remove vector sequence and contains a minimum of 100 bases of phred value 20 or above. For more details on library preparation and sequence analysis see <http://www.genome.clemson.edu/projects/barley>. To order this clone see <http://www.genome.clemson.edu/orders> see Close TJ, Wing R, Kleinof A, Wise R (2001) Genetically and physically anchored EST resources for barley genomics. Barley Genetics Newsletter 31:29-30. (<http://wheat.pw.usda.gov/g9pages/bgn/31/cover.html>)"

BASE COUNT 205 a 264 c 301 g 113 t
ORIGIN

Alignment Scores:

Pred. No.: 0.166 Length: 883
Score: 115.50 Matches: 71
Percent Similarity: 33.69% Conservative: 24
Best Local Similarity: 25.18% Mismatches: 115
Query Match: 4.79% Indels: 73
DB: 12 Gaps: 11

US-09-825-414-7 (1-486) x BG369810 (1-883)

QY 249 GYGLYLEUALAALAsnAlaGlyPheGlyAsnArgLeuSerValGlnSerArgAsp 268
 ||||| ||| ||| |||
Db 814 GGTGGTTTGGTGCCACGGCAAGCTTC-----788

QY 269 HisGln-----ArgGlyGlyAlaLeuValLeuGlyLeuLysAspLysGluProLysAla 286
 ||||| ||| ||| ||| ||| ||| ||| ||| |||
Db 787 CACAGGTGCTTCCGAAGTGTGTTCTCTCGCATTCGCAACGGGCAATCCTCATGGG 728

QY 287 GlnLeuSerGluGluAsnAsp-----TrrpleuGluAlaTrrLysAlaIleLysSerAla 304
 ||||| ||| ||| ||| ||| ||| ||| ||| |||
Db 727 CCGGTGTCTCTATATGCCACCGGCTTTGGGCTTCGGCTCGCCACGGCTGGCTTGGC 668

QY 305 SerTyrSerGlyAlaAlaLeuAsnAlaGlyLysArgMetAlaGlyLeuProLeuAspMet 324
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 667 TCTTGTGCTCGGCTCTCGGCGGAGACCTCT--TGCTGTGTTCATTGTTCTCGGTC 611

QY 325 AlaThrAspAlaMetGlyAlaValArgSerLeuValSerAlaSerSerLeuThrGlnAsn 344
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 610 TCGGCCGAGACGCTTCTTCTGTGTTCTGTTTCGGCTCA-----566

QY 345 GlyLeuAlaLeuAlaGlyGlyPheAlaGlyValGlyLysLeuGlnLumetalThrLys 364
 ||||| ||| ||| ||| ||| ||| ||| ||| |||
Db 565 -----GCCGGTGAT-----GCCTCACGCTCCACCACCT 539

QY 365 AsnIleThrAspProAlaThrLysAlaAlaValSerGlnLeuThrAsnLeu-----381
 ||||| ||| ||| ||| ||| ||| ||| ||| |||
Db 538 TCGGTCTCTGTCCCTCGGCGGAGGAGCTTCGGGCTCTTGTGTCGGTCACTGTGGGC 479

QY 382 -----AlaGlySerAlaAlaValPheAlaGlyTrrPthrThr-----AlaAlaLeu 396
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 478 GTGACAGGGGCGCTCTCAGCGGCGGCGGCTGGACAACTCTGTCTCGGCTCATGTC 419

QY 397 ThrThrAspProAlaValLys-----LysAla 405
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 418 GTCTCCGCTGGAGCTGCTCGGGCTTTTCGCCAGCTCGGTACAGGGCGGCGAGCAACT 359

QY 406 GluSerPheIleGlnAspThrValLysSerThrAlaSerSerThrThrGlyTrrValAla 425
 ||||| ||| ||| ||| ||| ||| ||| ||| |||
Db 358 TCCTCTTACAGGGCTCTCTCGGCGGAGCAACAGGGCGGCTCGACACATCAGCCGTTGCC 299

QY 426 AspGlnThrValLysLeuAla-----Lys 433
 ||||| ||| ||| ||| ||| ||| ||| ||| |||
Db 298 ACTTCGACTGTGTGACAGGGCGGCTCGACAATCTCCACTGCTGCCACCGTAGCCTCCTCG 239

QY 434 ThrValLysAspMetGlyGlyAlaIleThrHisThrGlyAlaSerLeuArgAsnThr 453
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 238 ACCGGAGCAGCGCGGCTCGACGACCTTCACCGCCAGCTAGTATCTTGACAAACGACG 179

QY 454 ValAsnAsnLeuArgGlnArgProAlaArgGluAlaAspIleGluGly-----471
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 178 ACGGCGGCTCGACAGCTTCCACCACCGCGGCGCTGGAGCCTCTCGACGGCATGGCG 119

QY 472 -----ThrAlaAlaSerProSerGluIleProPheArgPromet 484
 ||||| ||| ||| ||| ||| ||| ||| ||| |||
Db 118 TGGGTTCGGCGCTGACCTGACGCGCGCCATTGATGACAGATT-CCGAGTCGACCGGTC 60

QY 485 ArgSer 486
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 59 CGAACT 54

RESULT 6

AY103948/c 1792 bp mRNA linear HTC 25-MAY-2002
LOCUS AY103948
DEFINITION Zea mays PCO124731 mRNA sequence.
ACCESSION AY103948
VERSION AY103948.1 GI:21207026
KEYWORDS HTC.
SOURCE Zea mays.
ORGANISM Zea mays.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoideae; Andropogoneae; Zea.

REFERENCE

1 (bases 1 to 1792)
Hainey,C.F., Dolan,M., Miao,G.H., Vogel,J.M., Whitsitt,M.S.,
Arthur,L.W., Hanafey,M., Morgante,M. and Tingey,S.V.

TITLE

Maize Mapping Project/Dupont Consensus Sequences for Design of
Overgo Probes

JOURNAL

Unpublished (2002)
2 (bases 1 to 1792)

AUTHORS

Coe,E.C.
Submitted (25-APR-2002) Maize Mapping Project, University of
Missouri, Columbia, MO 65211, USA

JOURNAL

location/Qualifiers
1. 1792

FEATURES

source
/organism="Zea mays"
/db_xref="MaizeDB:637267"
/db_xref="taxon:4577"
/clone="PCO124731"
/clone_11b="Maize Mapping Project/Dupont Consensus
Library"

/note="this sequence is part of a project of EST
assemblies resulting from the application of public
contigs to seed Dupont contigs; this resource was
assembled by Dupont as part of a collaboration for the
overgo addressing of BACS in conjunction with the Maize
Mapping Project"

BASE COUNT 467 a 506 c 482 g 337 t
ORIGIN

Alignment Scores:

Pred. No.: 0.693 Length: 1792
Score: 114.00 Matches: 105
Percent Similarity: 32.95% Conservative: 37
Best Local Similarity: 24.36% Mismatches: 153
Query Match: 4.73% Indels: 136
DB: 11 Gaps: 20

US-09-825-414-7 (1-486) x AY103948 (1-1792)

QY 69 AlaAlaValAspValHisAsnAlaGlnIleThrAlaLeuIleGluThrArgAlaSerArg 88
 ||||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1636 GCAGCCATACAAAAGAGCTCAATGACAGAGCCATAATTACCAACATG-----1586

QY 89 LeuHisPheGluGluThrProAlaThrIleAlaAspThrPheAlaLysAlaGluLys 108
 ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1585 -----TCACCCAGCAGCGTCACTGACTACTATCAATGCTTACCGCA 1544

QY 109 LeuAspArgLeuAlaThrThrThrSer-----GlyAlaLeu 120
 ||||| ||| ||| ||| ||| ||| ||| ||| |||

Db 1543 TCCATCAGATTACAATGCGAATGCAAAATGTTTGTCTACATGAGATACACGGTGTGGA 1484
 QY 121 ArgAlaThrProPhe-----AlaMetaLaseSerLeuGlnTyrMetGlnProAlaIle 138
 Db 1483 AGGGGTGTCCGTTCCGGCTTCAGATGCACCTTCAGCTTGAGTGCCCGTTGTTCACGTTGT 1424
 QY 139 AsnLysGlyAspTrpLeuProAlaProLeuLysProLeuThrProLeuIle-----Ser 156
 Db 1423 TTCACGGGC---TGGCTGCTCTCCCTCGCTTCGCCCTTGTCCGACAGCGGTGTCTCTAGT 1367
 QY 157 GlyAlaLeuSerGlyAlaMetaspGlnValGlyThrLysMetMetLaspargAlaThrGly 176
 Db 1366 GGAGCGGCGCTGTGGCTCG-----GCTCCAGGC 1340
 QY 177 AspleuHisTyrLeuSerAlaSerProAspArgLeuHisAspAlaMetaLaseVal 196
 Db 1339 -----GCGGCATCAGTG 1328
 QY 197 LysArgHisSerProSerLeuAlaArgGlnValLeuAspThrGlyValAlaValGlnThr 216
 Db 1328 ----- 1328
 QY 217 TyrSerAlaArgAsnAlaValArgThrValLeuAlaProAlaLeuAlaSerArgProAla 236
 Db 1327 -----GCCGCAGCATCTGCCTGTGTCGCCGTGCTGCATCAGCGCGGCA 1283
 QY 237 ValGlnGlyAlaValAspLeuGlyValSerMetaLacGlyLeuAlaAlaAsnIacGly 256
 Db 1282 -----GCATCTGATTG---ATTTCAGGTGCTGATCAGCGGACGA---GCAGGC 1238
 QY 257 PheGlyAsnArgLeuLeuSerValGlnSerArgAspHisGlnArgGlyGlyAlaLeuVal 276
 Db 1237 TCGGCATCTGATTGGTTTCA-----GCCGCTGGAGCAGCAACA 1199
 QY 277 LeuGlyLeuLysAspLysGluProLysAlaGlnLeuSerGluGluAsnAspTrpLeuGlu 296
 Db 1198 GCATCAGCAGCAGAT----- 1184
 QY 297 AlaTyrLysAlaIleLysSerAlaSerTyrSerGlyAlaAlaLeuAsnIacGlyLysArg 316
 Db 1183 -----TCAGCAACTGGTTTGGTCTCGGGCAGCTGACAGCA----- 1148
 QY 317 MetaLacGlyLeuProLeuAspMetaLacThrAspAlaMetGlyAlaValArgSerLeuVal 336
 Db 1147 ---GCAGGCTCGGCATCTGATTGTTTCAGGCGCGCCGACGACGACATCAGCAGCA 1091
 QY 337 ---SerAlaSerSerLeuThrGlnAsnGlyLeuAlaLeuAlaGlyGlyPheAla----- 353
 Db 1090 GATTCAGCACTAGTTGGTCTCGGGCAGCTGAGCAGCAGCAGCAGGTTTCAGCACTGAT 1031
 QY 354 -----GlyValGlyLysLeuGlnLumetaLacThrLysAsnIleThrAspProAla 370
 Db 1030 TTGGTGTACAGCGCGCTGGAGCAGCAGCGCTCAGCACTGATTGTTGTCAGGCACTGGA 971
 QY 371 ThrLysAlaAlaValSerGlnLeuThrAsnLeu-----AlaGlySerAla 385
 Db 970 GCAGCAGCAGCAGCAGGCTCGGCATCTGTTGTTTCAGGCGCTGGGCGACGACAGCA 911
 QY 386 AlaValPheAlaGlyTyrThrThrAlaAlaLeuThrThrAspProAlaValLysLysAla 405
 Db 910 GCAGGCTTGGCATCTGTTTGGTTTCAGGCGCTGGGCGAGAAACAGCAGCAGGCTCGGCA 851
 QY 406 GluSerPheIleGlnAspThrValLysSerThrAlaSerSerThrThrGlyTyrValAla 425
 Db 850 ACTGATTTCGCTCAGGCACTGGGCGACGAAACAGCAGCGCTCGGACAGCAGCGGCT 791
 QY 426 AspGlnThrValLysLeuAlaLysThrValLysAspMetGlyGlyAlaIleThrHis 445
 Db 790 GAGTTCCTCAGTTGGTTG-----CTTACTGGAGCTGAGCCACAGCT 749
 QY 446 ThrGlyAlaSer---LeuArgAsnThrValAsnAsnLeuArgGlnArgProAlaArgGlu 464
 Db 748 TCAGGCTCATCTGGCTTGGCACTCTTCTTCTCTCC----- 713

QY 465 AlaAspIleGluGluGlyThrAlaAlaSer 475
 Db 712 TCCGATTGCCGACAGAGAGCTACACAGCTTCA 680

RESULT 7 4088 bp mRNA linear HTC 07-AUG-2002
 BC021514
 LOCUS
 DEFINITION
 Mus musculus, similar to Treacher Collins Franceschetti syndrome 1, homolog, clone IMAGE:5351783, mRNA.

ACCESSION BC021514
 VERSION BC021514
 KEYWORDS GI:18204412
 SOURCE HTC.

ORGANISM house mouse.

Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.

REFERENCE
 1 (bases 1 to 4088)
 Strausberg, R.
 Direct Submission

TITLE Submitted (14-JAN-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

REMARK NIH-MGC Project URL: http://mgc.nci.nih.gov
 Contact: MGC help desk

COMMENT
 Email: cgapbs-remail.nih.gov
 Tissue Procurement: Jeffrey Green M.D.
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NISC),
 Gaithersburg, Maryland;
 Web site: http://www.nisc.nih.gov/
 Contact: nisc_mgc@nih.gov

Akhter, N., Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B., Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S., Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghighi, P., Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Laric, P., Legaspi, R., Maduro, Q.L., Mastello, C., Maskeri, B., Mastrian, S.D., McCloskey, J.C., McDowell, J., Pearson, R., Stantripop, S., Thomas, P.J., Touchman, J.W., Tsurgeon, C., Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L., Young, A., Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: http://image.lnl.gov
 Series: IRAK Plate: 55 Row: 1 Column: 23

This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 6755741
 This clone has the following problem: frame shifted.

FEATURES
 Location/Qualifiers

source

1. 4088
 /organism="Mus musculus"
 /db_xref="taxon:10090"
 /map="FVB/N"
 /clone="IMAGE:5351783"
 /tissue_type="Mammary tumor, C3(1)-Tag model, Infiltrating ductal carcinoma, 5 month old virgin mouse."
 /clone_lib="NCI_CGAP_Mam6"
 /lab_host="DH10B"
 /note="Vector: pCMV-SPORT6"
 BASE COUNT 1100 a 1210 c 1210 g 568 t
 ORIGIN

Alignment Scores:

Pred. No.: 2.72 Length: 4088
 Score: 113.50 Matches: 124
 Percent Similarity: 32.40% Conservative: 73
 Best Local Similarity: 20.39% Mismatches: 225
 Query Match: 4.71% Indels: 186
 DB: 11 Gaps: 24

US-09-825-414-7 (1-486) x BC021514 (1-4088)

QY 11 ProValThrAla---ThrAspSerPheArgThrAlaSerAspAlaSerLeuAlaSerSer 29
 ||||| ||| ||||| ::: |||
Db 1186 CCAGTCACTGTGAATACTCTCTCAGGCCAAGCACTTCTGGAGAAGACCCCTGGGCCAGAGGT 1245
QY 30 SerValArgSerValSerSerAspGlnArgGluIleAsnAlaIleAlaAspTyrLeu 49
 ::: |||::: |||::: |||:::
Db 1246 ACCTCAGCCCCCCCCAAGAGTCAATCCCAAGAAAGGGCTCTGTGACGTCAACCCCTGAAAG 1305
QY 50 ThrAspHisValPheAlaAlaHisLysLeuProProAlaAspSerAlaAspGlyGlnAla 69
 ||| ||| ||| ||| |||:::
Db 1306 GCAAGGCTGTGGCAGCCCAAGCAGGCAAGAAACCAAGCAAGACAGTGTAG----- 1356
QY 70 AlaValAspValHisAsnAlaGlnIleThrAlaLeuIleGluThrArgAlaSerArgLeu 89
Db 1357 -----GAGTCAGAGAGTGAC-----GAGTCAGAGAGTGAC----- 1371
QY 90 HisPheGluGlyGluThrProAlaThrIleAlaAspThrPheAlaLysAlaGluLysLeu 109
Db 1372 -----AGTGGGAGACACACAGCTGCT----- 1392
QY 110 AsparGluAlaThrThrThrSerGlyAlaLeuArgAlaThrProPheAlaMetAlaSer 129
Db 1393 -----GCCACTCTGACACCAAGTCTCTGCCAAGGTGAACCTTTGGGGAAGAGCTCC 1443
QY 130 LeuLeuGlnTyrMetGlnPro-----AlaIleAsnLysGlyAspTyrPro 145
Db 1444 CAGGTCAGACCTGTCTTCCACCGTCACCCCGGGGTCTATCGGAAAAGGTGCCAACCTGCC 1503
QY 146 AlaProLeuLys----- 149
Db 1504 TGCCCTGGGAAGGTGGGGTCAAGAGCTCTCAAGGCTCCAATGTATAAGAAAGAATGTGC 1563
QY 150 -----ProLeuThrProLeu 154
Db 1564 TCGAGAGCAGCAGTGCAGAGCTGGACAGTGAAGGGCCTGGAGCCCAAGGCCAAAG 1623
QY 155 IleSerGlyAlaLeuSerGlyAlaMetAspGlnValGlyThrLysMet---MetAspArg 173
Db 1624 GCCTCCCTTGGCGTCCCTCAGAAAGGTGAGGCTGTGGCCACCCAGGTCACAGACTGACAGG 1683
QY 174 AlaThrGlyAspLeuHisTyrLeuSerAlaSerProAspArgLeuHisAspAlaMet--- 192
Db 1684 GGCMAAGGCCACTCAGGGAGCAGTGAAGTCTATCTGACAGCGAAGAGAGGAGCAGCACC 1743
QY 193 AlaAlaSerValLysArgHisSerProSerLeuAlaArgGlnVal----- 207
Db 1744 GCAGCCTGTGCTGCTCAGGCTAAGCCAGCTGTGAAAAGCAGATGAAGCTTCTCTAGG 1803
QY 208 -----LeuAspThrGlyValAlaValGlnThrTyrSerAlaArgAsnAla 222
Db 1804 AAAGCAGCGCTGCATCCGCAACAGAGCGACACCTCGTCCCATTTGTAGCGCAGAGCGG 1863
QY 223 Val-----ArgThrValLeuAlaProAlaLeuAla----- 232
Db 1864 GTGACCTCTTCAACCAAGCCTGTCTATCCCAAGCTGTGGCCAAAGGCAAGAGGTCAAGT 1923
QY 233 -----SerArgProAlaValGln 238
Db 1924 GTGACTCTTCCAGTGAAGTCTGAGTCAAGAGAGCTGCTCCAGACACCCCAAGGTACAG 1983
QY 239 GlyAlaValAspLeuGlyValSerMetAlaGlyLeuAlaAlaAsnAlaGlyPheGly 258
Db 1984 GGGAAAG-----TCTGGGGGCAAGGGCCTCCAAGGAAAGCTGCCTTGGGG 2028
QY 259 AsnArgLeuLeuSerValGlnSerArgAspHisGlnArgGlyGlyAlaLeuValLeuGly 278
Db 2029 CAAGGGGTGGCCCCAGTGCACACT-----CAGAAGACAGGGCCTTCGGTCAAAAGCT 2079
QY 279 LeuLysAspLysGluProLysAla-----GlnLeuSerGluGluAsnAspTyr 294
Db 2080 ATGGCTCAGGAAGACTCAGAGAGCCTCGAGGAGACTCCAGCAGCGAAGAGAGATGAG 2139

QY 295 LeuGluAlaTyr-----LysAlaIleLysSerAla 304
Db 2140 ACCCCAGCACAGGCCACGCCCTTGGGAGACTTCCTCAGGCCAAAGCCAAACCACTCCC 2199
QY 305 SerTyrSerGlyAlaAlaLeuAsnAlaGlyLysArgMetAlaGly----- 319
Db 2200 ACTAAGACACCTCCAGCGCTCTGCATCTGGAAGAAAGCTGTGGCTGCTCCAACCAAGGAAA 2259
QY 320 LeuProLeuAspMetAlaThrAspAlaMetGlyAlaValArgSerLeuValSerAla--- 338
Db 2260 CCACCTGTCCGACACAGCACCGCTCTGTGCAAGGGGCCAAGCGGTCTGTGCCAGCCGGGA 2319
QY 339 -----SerSerLeuThrGlnAsnGlyLeuAlaLeuAlaGlyGlyPheAlaGlyVal 355
Db 2320 AAAGCAGGGGGCCCAAGCAACCCAA-----GCCAGAAAGGTCCTGGCTGGCACA 2370
QY 356 GlyLysLeuGlnGluMetAlaThrLysAsnIleThrAsp-----ProAla 370
Db 2371 GGGGAGACTCAGAGACAGCAGTAAAGAGAGTCTGACAGTGAAGAAGACAGCCAGCC 2430
QY 371 ThrLysAlaAlaValSerGlnLeuThrAsnLeuAlaGlySerAlaAlaVal----- 387
Db 2431 CAGATAAAACCTGTGGGGAAGACCTCTCAGGTCAGAGCTGCCTCAGACCCCTGCCAAGGAG 2490
QY 388 -----PheAlaGlyTyrThrThrAlaAlaLeuThr--- 397
Db 2491 TCTCCTAAAAAGAGGCCCATCCAGGAACCCCGGCAAGACGGGATCTTCAGCTACCCAG 2550
QY 398 -----ThrAspPro 400
Db 2551 GCCCAGCCAGGGAAGACAGAGGACTCAGACAGCAGTAGTGAAGAGTCTGACAGTGAACACA 2610
QY 401 AlaValLysLysAlaGluSerPheIleGlnAspThrValLysSerThrAlaSerSerThr 420
Db 2611 GAGATGCCATCAGCCAGGATGCTATCTCCACGCTGCCAGAGGCAAGGCTCAGGTCCA 2670
QY 421 ThrGlyTyrValAlaAspGlnThrValLysLeuAlaLysThrValLysAsp----- 437
Db 2671 GCCTCC-----CCAGAGAAGAGCATAGAAGGCTCCTCAGAGACAGTGAAGATCTG 2724
QY 438 MetGlyGlyGluAlaIleThrHisThrGlyAlaSerLeuArgAsnThrValAsnLeu 457
Db 2725 CCCTCCGGCCAGCGCATTAATCCCTCCAGTTCTGTG-----AACCGT 2769
QY 458 ArgGlnArgProAlaArgGluAlaAspIleGluGluGlyThrAlaAla-SerProse 477
Db 2770 AACAGTAGTCCAGCTGTCCAGCTCTTACCCCAAGAGAGTCCAGGCTGTGAACACCACA 2829
QY 477 rGluIleProPheArgPromet 484
Db 2830 AAGAAGCCTCAGCAGCACCACTG 2851
RESULT 8
AY107104
LOCUS 2012 bp mRNA linear HTC 25-MAY-2002
DEFINITION Zea mays PCO148976 mRNA sequence.
ACCESSION AY107104
VERSION AY107104.1 GI:21210182
KEYWORDS HTC.
SOURCE Zea mays.
ORGANISM Zea mays.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 2012)
AUTHORS Hainey,C.F., Dolan,M., Miao,G.H., Vogel,J.M., Whitsitt,M.S.,
Arthur,L.W., Hanafey,M., Morgante,M. and Tingey,S.V.
TITLE Maize Mapping Project/Dupont Consensus Sequences for Design of
Overgo Probes
JOURNAL Unpublished (2002)
REFERENCE 2 (bases 1 to 2012)
AUTHORS Coe,E.C.
TITLE Direct Submission

BASE COUNT 173 a 268 c 371 g 162 t 1 others
ORIGIN
Alignment Scores:
Pred. No.: 0.502 Length: 975
Score: 111.50 Matches: 83
Percent Similarity: 37.30% Conservative: 55
Best Local Similarity: 22.43% Mismatches: 127
Query Match: 4.62% Indels: 105
DB: 9 Gaps: 16
US-09-825-414-7 (1-486) x AL545051 (1-975)

QY 53 ValPheAlaAlaHisLysLeuProProAlaAspSerAlaAspGlyGlnAla--AlaVal 71
Db 60 ATAACGGCGCGAGCGGCGAAACCTCGAGCTCGAGGGCGGCGCAAGCGCGGCTTG 119
QY 72 AspValHisAsnAlaGlnIleThrAlaLeu-----IleGluThrArg 85
Db 120 GAGATGGAGAATTCAGATTGTAGAGCTGTCATCGCGGCGCTCATGTGCAGACGAGT 179
QY 86 AlaSerArgLeuHisPheGluGlyGluThrProAlaThrIleAlaAsp----- 101
Db 180 GAGTCGGGCGCTGCGCGCCCACTTTGAGGCGCTTGGGACTCTGACGGAGCTGCGGTG 239
QY 102 -----ThrPheAlaLysAlaGlu 107
Db 240 GTGAATCCCGACAGCAAGCGCTCCGTTGCTTGGCTTCGTGACCTACTCCAAATGGAG 299
QY 108 LysLeuAspArgLeuAlaThrThrThrSerGlyAlaLeuArgAlaThrProPheAlaMet 127
Db 300 GAGCGCGAC-----GCCGCCATGCCCGCCCTCGCCCATGCCGTG 338
QY 128 AlaSerLeuLeuGlnTyrMetGlnProAlaIleAsnLysGlyAspTrpLeu--ProAla 146
Db 339 GACGGCAACACTGTGAGCTGAAGCGGGGTGTCCCGGAGAGATTGCGCGCGCGCGGT 398
QY 147 ProLeuLysProLeuThrProLeuIleSerGlyAlaLeuSerGlyAlaMetAspGlnVal 166
Db 399 GCCCAGCCAGGTTAAGAACTTTGTGTCGAGGCGCTTAAAGGA----- 443
QY 167 GlyThrLysMetLeuAspArgAlaThrGlyAspLeu-----HisTyr-LeuSerAlaSe 184
Db 444 -----GACGTGGCTGAGGGCGAGCGACTGATCGACTTCTCGCAGTTGGC 488
QY 184 rProAspArgLeuHisAspAlaMetAlaAlaSerValLysArgHisSerProSerLeuAl 204
Db 489 ACCGTGAAAGCGCCGAGATTATTGCCGACAAGC-----AGTCCGGC 530
QY 204 aArgGlnValLeuAspThrGlyValAlaValAlaGlnThrTyrSerAlaArgAsnAlaValAr 224
Db 531 AAGAAAGCGT--GATTCGCTTCGTGTATTTCAGAAATCAGACGCGGAG----- 579
QY 224 gThrValLeuAlaProAlaLeuAlaSerArgProAlaValGlnGlyAlaVal-AspLeuG 244
Db 580 -----ACAAGCGCGCGGTGTCAGTTCCATCCGATTTCAG 614
QY 244 LysValSerMetAlaGlyGlyLeuAlaAlaAsnAlaGly-----PheGly- 258
Db 615 GGCCATTCGCGTGGAGGTGAAGAAAGCAGTCCCAAGGAGATATCTACTCCGGTGGGGGT 674
QY 259 --AsnArgLeuLeuSerValGlnSerArgAspHisGlnArgGlyGlyAlaLeuValLeuG 278
Db 675 GGAGGGCGGCTCCGATCTCCCGGGGGCGCGGAGCGCGGGGGGGGGGGGGGTGCG 734
QY 278 LysLeuLysAspLysGluProLysAlaGlnLeuSerGluGluAsnAspTrpLeuGluAlaT 298
Db 735 AGACCAGAACGGCTTTCCAAGGGCGCG----- 763
QY 298 yLysAlaIleLysSerAlaSerTyrSerGlyAlaAlaLeuAsnAlaGlyLysArgMetAl 318
Db 764 --GCGGCGGTTACAACAGCTACGCTGTTACGGCGCGCGGAGGCGCG----- 811

QY 318 IagLysLeuProLeuAspMetAlaThrAspAlaMetGlyAlaValArgSerLeuValSera 338
Db 812 -----GCTACAAATGCTTACGAGGCGGCGCGGCTTCCTCCTACGCTGGAG 860
QY 338 IaserSerLeuThrGlnAsnGlyLeuAlaLeuAlaGlyGlyPheAlaGlyValGlyLysL 358
Db 861 CGACTACGCTAACGGCTTCGCGCGCTTCG----- 889
QY 358 euGlnGluMetAlaThrLysAsnIleThrAspProAla-----ThrLysAlaAlay 375
Db 890 -----GCAGCTACAGCCAGCATCAGTCTCTATGGGCCCATGAGAGCGGCGCG 941
QY 375 aLSerGlnLeuThrAsnLeuAlaGly 383
Db 942 CGGCGCGGCTGAGGAGCAGTACTGGG 967

RESULT 10
BM044579/c 948 bp mRNA linear EST 07-NOV-2001
LOCUS 603622231F1 NIH_MGC_40 Homo sapiens cDNA clone IMAGE:5447983 5',
DEFINITION mRNA sequence.
ACCESSION BM044579
VERSION BM044579.1 GI:16773846
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 948)
AUTHORS NIH-MGC http://mgc.ncl.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabs-remail.nih.gov
Tissue Procurement: DCTD/DTP
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
plate: L1CM1932 row: m column: 08
High quality sequence stop: 744.
Location/Qualifiers

FEATURES
source
1..948
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5447983"
/clone_1ib="NIH_MGC_40"
/tissue_type="carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: prostate; Vector: pOTB7; Site_1: XhoI;
Site_2: EcoRI; cDNA made by oligo-dT priming.
directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCACGAG(G). Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies).
Note: this is a NIH_MGC library."

BASE COUNT 179 a 274 c 273 g 221 t 1 others
ORIGIN

Alignment Scores:
Pred. No.: 0.542 Length: 948
Score: 111.00 Matches: 81
Percent Similarity: 40.00% Conservative: 37
Best Local Similarity: 27.46% Mismatches: 102
Query Match: 4.60% Indels: 75
DB: 13 Gaps: 15

US-09-825-414-7 (1-486) x BM044579 (1-948)
QY 218 SerAlaArgAsnAlaValArgThrValLeuAlaProAlaLeuAla----- 232

OY	376	SergInLeuthrAsnleuAlaGlySerAlaAlaValPheAlaGlyTrpThrAlaAla	395
	:::	:::	
Db	499	GCAGCAACAGCAACAGCAACAGCAACAGCAGCAACAACAGCAACA--ACAACAGCAGCA	555
OY	396	LeuthrThraspproAlaValLysLysAlaGluSerPheIleGlnAspThrValLysSer	415
			:::
Db	556	ACAACAGCAGATGCAGCAACAGCATGCA-----ACAGCAACAGCA	597
OY	416	ThrAlaSerSerThrThrGlyTyrValAlaAspGlnThrValLysLeuAlalystThrVal	435
	:::		
Db	598	GCAGCCGCACACACACTGCAGCAGCAGCAGCGCCGCCAGCAGCAGCGCCGCCA	657
OY	436	LysAspmetGlyGlyAlaIleThrHisThrGlyAlaSerLeuArgAsnthr	453
			:::
Db	658	GCAACAACAACAACATCTGCAGCAACAGCAGCAGCACACGACCAACAACAACACA	711
RESULT 12	CNS076CP/c	888 bp	DNA linear GSS 07-JUL-2001
LOCUS	T7 end of clone BB0AA002G01 of library BB0AA from strain CBS 4732		
DEFINITION	of Pichia angusta, genomic survey sequence.		
ACCESSION	AL431231		
VERSION	AL431231.1	GI:12214643	
KEYWORDS	GSS.		
SOURCE	Pichia angusta.		
ORGANISM	Pichia angusta.		
REFERENCE	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;		
AUTHORS	Saccharomycetales; Saccharomycetaceae; Pichia.		
	1 (bases 1 to 888)		
	Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G.,		
	Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,		
	de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Lorente,B.,		
	Malpertuy,A., Neuveglise,C., Ozier-Kalogeropoulos,O., Potier,S.,		
	Saurin,W., Tekala,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,		
	Wincker,P. and Weissenbach,J.		
TITLE	Genomic exploration of the hemiascomycetous yeasts: 1. A set of		
JOURNAL	yeast species for molecular evolution studies		
MEDLINE	FEBS Lett. 487 (1), 3-12 (2000)		
PUBMED	20584711		
REFERENCE	11152876		
AUTHORS	2 (bases 1 to 888)		
	Blandin,G., Llorente,B., Malpertuy,A., Wincker,P., Artiguenave,F.		
	and Dujon,B.		
TITLE	Genomic exploration of the hemiascomycetous yeasts: 13. Pichia		
JOURNAL	angusta		
MEDLINE	FEBS Lett. 487 (1), 76-81 (2000)		
PUBMED	20584723		
REFERENCE	11152888		
AUTHORS	3 (bases 1 to 888)		
	Genoscope.		
TITLE	Direct Submission		
JOURNAL	Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage,		
MEDLINE	2 rue Gaston Cremlieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :		
PUBMED	segref@genoscope.cns.fr - Web : www.genoscope.cns.fr)		
REFERENCE	This GSS is part of a random genomic sequencing program of thirteen		
AUTHORS	yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces		
TITLE	exiguus, Saccharomyces servazii, Zygosaccharomyces rouxii,		
JOURNAL	Saccharomyces kluyveri, Kluyveromyces marxianus var. marxianus, Pichia		
MEDLINE	lactis var. lactis,Kluyveromyces thermotolerans, Kluyveromyces		
PUBMED	angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,		
REFERENCE	Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to		
AUTHORS	5 kb were prepared and both extremities were sequenced. See		
TITLE	keywords for description of this sequence and for the sequence of		
JOURNAL	the other extremity of this insert.		
MEDLINE	Location/Qualifiers		
PUBMED	1. .888		
REFERENCE	/organism="Pichia angusta"		
AUTHORS	/strain="CBS 4732"		
TITLE	/db_xref="taxon:4905"		
JOURNAL	/clone="BB0AA002G01"		
MEDLINE	/clone_lib="BB0AA"		
PUBMED	/note="end : T7"		
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			

[illegible]

```

Db      72  AGTTTATCCTTG-----CCTAGTGCACCAATCAGCTCTGTGTCAAACAGCA 28
              ::|||  |||  |||  |||  |||
QY      474  Alaser 475
              |||
Db      27  ATTCA 22
              |||

RESULT 13
LOCUS    BQ424977/c 860 bp mRNA linear EST 23-MAY-2002
DEFINITION AGENCOURT_7909558 NIH_MGC_71 Homo sapiens cDNA clone IMAGE:6156315
5', mRNA sequence.
ACCESSION BQ424977
VERSION   BQ424977.1 GI:21120292
KEYWORDS  EST.
SOURCE    human.
ORGANISM  Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 860)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
          Email: cgapbs-r@mail.nih.gov
          Tissue Procurement: ATCC
          cDNA Library Preparation: Life Technologies, Inc.
          cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
          DNA Sequencing by: Agencourt Bioscience Corporation
          Clone distribution: MGC clone distribution information can be
          found through the I.M.A.G.E. Consortium/LLNL at:
          http://image.llnl.gov
          Plate: LHAM13500 row: g column: 04
          High quality sequence stop: 618.
FEATURES
    source
        1..860
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="IMAGE:6156315"
            /clone_lib="NIH_MGC_71"
            /tissue_type="leiomyosarcoma"
            /lab_host="DH10B (phage-resistant)"
            /note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: NotI;
            Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
            Average insert size 2.1 kb."
BASE COUNT 217 a 274 c 247 g 120 t 2 others
ORIGIN
Alignment Scores:
Pred. No.: 0.595 Length: 860
Score: 110.00 Matches: 70
Percent Similarity: 36.84% Conservative: 28
Best Local Similarity: 26.32% Mismatches: 99
Query Match: 4.56% Indels: 69
DB: 14 Gaps: 9
US-09-825-414-7 (1-486) x BQ424977 (1-860)

QY      137  AlaiLeasnlGlyAspTrpLeuProAlaProLeuYsProLeuThrProLeuIleSer 156
          |||::: ||||| ||||| ::||| ||||| |||||
Db      702  GCTCTGACAGAATGGTGACCTTCAGGAGACCCCGCGGCCCATACCCCTGGCATCTCT 643
          ||||| ::|: |||||
QY      157  GLYAlaleuSerGly-AlaMetaspGlnValGly----- 167
          ||||| ::|: |||||
Db      642  CTCGCTCTGTTTTCACATCTCTCCACACATCTAGGCCACAGCCGACTCATTTCTGAATCTCT 583
          ||||| ::|: |||||
QY      168  -----ThrLysMetMetaspArgalaThrGlyAspLeuHisTyrLe 181
          ::|: |||||
Db      582  TCCATGCTGACTTGGGTGCCACACAGGCTCTGGCAAGGCTTTTCCATGCGCTCCACACAGCT 523
          ::|: |||||
QY      181  uSerAlaserProAspArgLeuHisAspAlaMetalaAlaserValLysArgHisSerPr 201
          | ||||| ::|||::|
Db      522  T-----CTACTGCATFCAG--GTGGCATCTCCAGCCCGCAGACATGAACC 481
          ||||| ::|||::|

```

QY	201	oserleu-----AlaargcInValleuaspthrGlyValAlaValGIInthrTyrSeral	219
Db	480	ACAACTGGTGTGTTCCTGTGGAGCCAGCAACGACACACTCGCGGCTGGGTGTGAGGCTGTCA	421
QY	219	aArgAsnAlaValArgThrValIleuAlaProAlaLeuAlaSerArgProAlaValGIInGl	239
Db	420	TGGGTCTGCTCTGGTGCGCTGTACTGCACACCTGCCGTGGAGCATGCTCCTGTG-----	369
QY	239	yAlaValasPleunGlyValSerMetAlagLygLyLeuAlaAlaAsnAlagLypheGLyas	259
Db	368	-GCTGGACGCTTGGCTGTGGCTGTGTGTGGACCCTGTGGCTGCACCTGCTTCTGTG-----	315
QY	259	nArgLeuLeuSerValGIInSerArgAspHisGIInArgGIyGLyAlaLeuValLeuGlyLe	279
Db	314	-----GCTCTG-----	309
QY	279	uLysAsplySGluProLysAlaGIInLeuSerGIuGIInasnsPTrrPleunGIAlaTyrLy	299
Db	308	-----CCFCTGCT-----	300
QY	299	sAlaIleLysSerAlaSerTyrSerGIyAlaAlaLeuAsnAlaGIyLysArgMetAlaGI	319
Db	299	-GCAGCTGTGGCTGCACCTGCTTCTGTGTGCAGACACTAAGTCTCTGCAGAGAGTCTGTGTT	241
QY	319	yLeuProLeuAspMetAlaThrAspAlaMetGIyAlaValaLarSerLeuValSerAlase	339
Db	240	TGCGCCTGCTTCTGCAGCTTTGGCTGCACCTGGGTGTGTGTGATGTATCCGCGCTGA	181
QY	339	rSerLeuThrGIInasnGIyLeuAlaLeuAlaGIyGIyPheAlaGIyValaGIyLysIeuGI	359
Db	180	GTCCTGTGACTGCACCTGGGCGCTGCACCTGCAGAGACCCGTGGCTGAATC-----	132
QY	359	nGIuMetAlaThrLysAsnIleThrAspProAlaThrLysAlaAlaValSerGIInLeuTh	379
Db	131	---GTGGCAGCACCTTGGGCTTCCAGGGCCTCAGGCAGCAGGCTGTGTGTGTGTGTGT	76
QY	379	rAsnLeuAlaGIySer	384
Db	75	TTCGGTACTGTCAATCC	60
RESULT 14			
LOCUS	BU027919	670 bp	mRNA
DEFINITION	QHGM04.yg.ab1 QH_EFGHJ sunflower RHA280 Helianthus annuus cDNA clone QHGM04, mRNA sequence.		
ACCESSION	BU027919		
VERSION	BU027919.1	GI:22463439	
KEYWORDS	EST.		
SOURCE	common sunflower.		
ORGANISM	Helianthus annuus		
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; euasterids II; Asterales; Asteraceae; Asteroideae; Heliantheae; Helianthus.		
AUTHORS	1 (bases 1 to 670) Kozik,A., Michelmore,R.W., Knapp,S., Matvienko,M., Riesberg,L., Lin,H., van Damme,M., Lavelle,D., Chevalier,P., Ziegler,J., Ellison ,P., Kolman,J., Slabaugh,M.S., Livingston,K., Zhou,Y., Lai,Z., Church,S., Jackson,L. and Bradford,K. lettuce and Sunflower ESTs from the Compositae Genome Project http://compgenomics.ucdavis.edu/ Unpublished (2002)		
TITLE	Contact: Alexander Kozik [R.W.Michelmore] Department of Vegetable Crops, R.W.Michelmore lab University of California at Davis (UCD) Asmundson Hall, UCD, Davis, CA 95616, USA Tel: 1-(530)-742-1742 Fax: 1-(530)-752-9659		
JOURNAL COMMENT	Email: akozikeatgc.org [michelmore@vegmail.ucdavis.edu] belongs to contig QH_CA_Contig3101, see http://cgpdb.ucdavis.edu/ for details.		

QY	6	ArgValGlnGlnProProValThrAlaThrAspSerPheArgThrAlaSerAlaSer	25
		::::::::::	
Db	74	CAATTGAAGCAAGACCCCAAGACAGTGAATCTGGGTCTTAATCCTGAGTCTGCCAGC	133
QY	26	LeuAlaSerSerSerValArg-----SerValSerSerAsp	37
		:::	
Db	134	CTGACATGGGATTCTCTCCCAAGACCAAGGTCTCCAGATGACACACGAAAGCAGTGAT	193
QY	38	GlnGlnArgGluIleAsnAlaIleAlaAspTyr-----LeuThrAspHisValPhe	54
		:::	
Db	194	CCAGGCAGGACAGCAGTTGTCTCCCTACAGACATAGAATGAGGCAAAAGCACCAT--TTT	250
QY	55	AlaAlaHisLysLeuProProAlaAspSerAlaAspGlyGlnAlaAlaValAspValHis	74
Db	251	AACTTTACTGGCCTCTCTCCCTGCACCAAGCTGCCCTCAGGACTGAATAACC-----	298
QY	75	AsnAlaGlnIleThrAlaLeuIleGluThrArgAlaSerArgLeuHisPheGluGlyGlu	94
		::: :::	
Db	299	TCTCCTTCTCGGGGAGGCGCTCTACACTAAGCGGTCTCCCATGAACCTTCCCCCAGCAA	358
QY	95	ThrProAlaThrIleAlaAspThrPheAlaLysAlaGluLysLeuAspArgLeuAla---	113
		:::	
Db	359	GGGAAAAGTTGAATGGGAT-----GTGAATGTTAATGGCTTATCTACTGTATCTCAC	412
QY	114	ThrThrThrSerGlyAlaLeuArgAlaThrProPheAlaMetAlaSer-----	129
Db	413	ACTACTACTTCAGGAGATTGAACTCTGCTCCACACTCCTGTAGCACCTCACACCTCCAT	472
QY	130	-----LeuLeuGlnTyrMetGln---ProAlaIleAsnLys	140
Db	473	CACCCCTAAGCTGGCCTACGACTGTCTTTGGAACTACTCACAAGTAAACCATCTGCCAATCCT	532
QY	141	GlyAspTyrPleu-----ProAlaProLeuLysProLeu	151
		:::	
Db	533	GGCAACAACCTCAAGACCCACCCCTCTTCTTCTCAGTTCCCTGGGGGACAATACCCTC	592
QY	152	ThrProLeuIleSerGlyAla-----	158
		:::	
Db	593	AACGGTATCCTTGGGGCAACCGACAACTTCATCCCCCAAGTCAACAACATAATCTTGCA	652
QY	159	-----LeuSerGlyAlaMetAspGlnValGlyThrLysMetMet	171
Db	653	GCTGGAGCCAAAGATTTCTGGGCCAATGGTACCAGAGTCCCATGGGGCTTAACCTTCGAT	712
QY	172	AspArgAlaThrGlyAspLeuHisTyrLeuSerAlaSerProAspArgLeuHisAspAla	191
		:::	
Db	713	TCACAGGAAGTGTATGAT-----TCTTTTCTGATCAGAAATTTTGAGGTG	757
QY	192	MetAla-----AlaSerValLysArgHisSerProSerLeuAlaArg	205
		:::	
Db	758	ATGCCCAATGAGACCCCAAGTTTTCACCTCCCTCAGACTTCTCCAATGTG-----	811
QY	206	GlnValLeuAspThrGlyValAlaValGlnThrTyrSerAlaArgAsnAlaValArgThr	225
		:::	
Db	812	-----GGGTAGTATCCAGACCTTTCACCTTCCCAGGATGTAAGCAGT	856
QY	226	ValLeuAlaProAlaLeuAlaSerArgProAlaValGlnGlyAlaVal--AspLeuGly	244
		:::	
Db	857	GACATCCATCCTGATGAAGACGAGAAAGAGAGCTGACTTCAGTTGTGGCAGAAATGGC	916
QY	245	ValSerMetAlaGlyGlyLeuAlaAlaAsn-----AlaGly	256
		:::	
Db	917	ACTGGCTTGGTAGGCAGCCTGAGAGCTGGAGAGAGCAGCAGAACTAAAGATGTGTGGC	976
QY	257	PheGlyAsnArgLeuLeuSerValGlnSerArgAspHisGlnArgGlyGlyAlaLeuVal	276
		:::	
Db	977	TACAATGGTCTGTCTCTCCCTGTGGAGTCTTTA--CACCAAGAAGTCTCCGCTCGGTGTC	1033
QY	277	-----LeuGlyLeuLysAspLys---GluPro	284
Db	1034	CCTGATCCCAACAGTGAGCTGTCTAGATGATCCTTCACATCTTCTCCTGATCAACTGGAAGAC	1093
QY	285	LysAlaGlnLeuSerGluGluAsnAspTyrPleuGlnAlaTyrLysAlaIleLysSerAla	304

Db	1094	ACTCCAATTTCTCAGTGA				:::	:::::	:::	
QY	305	SerTyrSerGlyAla	-----	AlaLeuAsnAlaGlyLysArgMetAlaGly	-----				319
Db	1148	CCAGTAGAGTGGCAGCTCTTATGGTATAGATGATGCGGAGCTGATGGGTGCAGAGACAAG		:::::	:::::	:::			1207
QY	320	LeuProLeuAsp	-----		-----				323
Db	1208	TTGCCTCTGGAGGGCAACCCCTGTGATCTCTGCCCCCTGCATTTGCCCTGCTCTCAGTAATGCT		:::::					1267
QY	324	-----	-----	MetAlaThrAspAlaMetGlyAlaValArgSerLeuValSerAla	-----				338
Db	1268	AATGCCTTCAGTCTCTCCCTGGCAGACGACAGCCAGACATCAGCCCTCCATCTTTGTGACGCCCT		:::::	:::::	:::			1327
QY	339	SerSer	-----	-----	LeuThrGlnAsnGlyLeuAlaLeuAlaGly	-----			350
Db	1328	ACCTCCCCACCTGTCTTTAGGGGAGTCTGTCTTGCAGATTAACAGCTTTGAGCTG							1381
QY	351	GlyPheAlaGlyValGlyLysLeuGlnGluMetAlaThrLysAsn	-----		-----				365
Db	1382	AACAGTTGCAGTGACTCTGAACAGAGAATAAGACCACCTTCAAACTTCCAACGT		:::::	:::::		:::::		1441
QY	366	-----	-----	-----	-----	ProAlaThrLys	-----		372
Db	1442	CCCCTGACTGAGCCAGCTCCTGACCAAGCCACCTAGTACTCACTACATCCAGCAGTTCA							1501
QY	373	AlaAlaValSerGlnLeuThrAsnLeuAlaGlySerAlaAlaValPheAlaGlyTrpThr	-----		-----				392
Db	1502	CCAACAGCCTCCCCAGCAGCCTCCTTGACAGCATCTGCAGAAATCTCTCCAGCTGTCTCT		:::::		:::::	:::::		1561
QY	393	ThrAlaAlaLeuThrThrAspPro	-----	-----	AlaValLysLysAlaGluSer	-----			407
Db	1562	CCAGTAGCATCTCGCCTGTCCCTCTGGAAGCTTTGTGAGCAAGTCTCTCCAGCTTCTCTCA		:::::					1621
QY	408	PheIleGlnAspThrValLysSerThrAlaSerSerThrThrGlyTyr	-----		-----				423
Db	1622	CCTGCTCTGCCAGCCATCTCTTGGAAGCCTGTATGACAACTCCAGTAACCTCTCTCAA							1681
QY	424	-----	-----	ValAlaAspGlnThrValLysLeuAlaLysThrVal	-----				435
Db	1682	GGTTCCCTGAACTTCTCCAGCAGCAGCTGCCTTCAGACTGTCTCCCCAGCAAGG							1735
QY	436	LysAspMetGlyGlyAlaIleThrHisThrGlyAlaSerLeuArgAsnThrValAsn	-----		-----				455
Db	1736	-----	-----	-----	-----	AAA	-----		1738
QY	456	AsnLeuArgGlnArgProAlaArgGluAlaAspIleGluGlu	-----	GlyGlyThrAla	-----				473
Db	1739	AATGTCAGCAGTCTCTCTTAAAGCGCGTGTGATCGAGAGAGACGACTGGAGGACAGTT							1798
QY	474	Ala 474							
Db	1799	GCA 1801							

Search completed: January 31, 2003, 07:16:15
Job time : 1595.38 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: January 31, 2003, 04:12:39 ; Search time 54.444 Seconds
(without alignments)
2737.582 Million cell updates/sec

Title: US-09-825-414-7

Perfect score: 2412

Sequence: 1 MHINRRVQGPVTATDSFRT.....IEEGTAASPEIFRPMPRS 486

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 441362 segs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool/US09825414/runat_28012003_161742_26114/app_query.fasta_1.1294
-DB=Issued_Patents_NA -QFMT=fastap -SUFFIX=p2n.rni -MINMATCH=0.1 -LOOPL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=biosum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09825414_@CGN_1_1_40_@runat_28012003_161742_26114 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued_Patents_NA:*

1: /cgn2_6/ptodata/1/ina/5A.COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B.COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A.COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B.COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS.COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result	Score	Query Match	Length	DB	ID	Description
-----8-----						
C 1	168	7.0	4403765	4	US-09-103-840A-2	Sequence 2, Appli
C 2	164	6.8	4411529	4	US-09-103-840A-1	Sequence 1, Appli
3	147.5	6.1	6744	1	US-08-119-125A-2	Sequence 2, Appli
4	143.5	5.9	46819	4	US-09-453-702B-72	Sequence 72, Appli
5	138.5	5.7	4411529	4	US-09-103-840A-1	Sequence 1, Appli
6	137.5	5.7	32768	4	US-08-961-527-71	Sequence 71, Appli
7	135.5	5.6	2550	4	US-08-961-527-364	Sequence 364, App
8	133.5	5.5	5036	4	US-09-177-349-2	Sequence 2, Appli
9	132.5	5.5	4403765	4	US-09-103-840A-2	Sequence 2, Appli
10	130.5	5.4	44377	2	US-08-804-227C-7	Sequence 7, Appli
11	130.5	5.4	44377	2	US-08-804-198-1	Sequence 1, Appli
12	129	5.3	4483	4	US-08-961-527-363	Sequence 363, App

C 13	128.5	5.3	38506	3	US-09-320-878-19	Sequence 19, Appli
14	128	5.3	16836	4	US-09-147-236-1	Sequence 1, Appli
15	128	5.3	16836	4	US-09-147-236-10	Sequence 10, Appli
16	127.5	5.3	3106	2	US-08-591-079-5	Sequence 5, Appli
C 17	127.5	5.3	11220	4	US-09-105-537-32	Sequence 32, Appli
18	127.5	5.3	15872	4	US-09-105-537-1	Sequence 1, Appli
C 19	127.5	5.3	36778	4	US-09-105-537-5	Sequence 5, Appli
20	127.5	5.3	50341	1	US-08-247-901C-1	Sequence 1, Appli
21	127.5	5.3	50341	2	US-09-075-904-1	Sequence 1, Appli
22	127.5	5.3	52297	4	US-09-426-436-1	Sequence 1, Appli
23	127.5	5.3	52297	4	US-08-705-557-1	Sequence 6, Appli
24	127	5.3	3251	4	US-09-085-199B-6	Sequence 1, Appli
25	127	5.3	3300	1	US-08-194-290-6	Sequence 6, Appli
26	127	5.3	3300	2	US-08-614-377A-6	Sequence 6, Appli
27	127	5.3	3300	4	US-09-142-648B-6	Sequence 6, Appli
28	127	5.3	9542	4	US-08-968-685A-9	Sequence 9, Appli
29	126.5	5.2	2406	1	US-07-689-008-4	Sequence 257, App
30	126.5	5.2	4977	4	US-09-071-035-257	Sequence 261, App
31	126.5	5.2	4977	4	US-09-071-035-261	Sequence 265, App
32	126.5	5.2	4977	4	US-09-071-035-265	Sequence 1, Appli
33	126.5	5.2	9540	1	US-07-689-008-1	Sequence 367, App
34	126	5.2	1702	4	US-08-961-527-367	Sequence 10, Appli
35	125.5	5.2	3979	4	US-09-085-199B-10	Sequence 3, Appli
C 36	123.5	5.1	8078	4	US-09-702-251-3	Sequence 1, Appli
C 37	123	5.1	6530	2	US-08-146-930-1	Sequence 1, Appli
C 38	123	5.1	6530	3	US-08-458-240-1	Sequence 1, Appli
C 39	123	5.1	6530	5	PCT-US93-03993-1	Sequence 6, Appli
40	122.5	5.1	28958	1	US-08-258-261B-6	Sequence 6, Appli
41	122.5	5.1	28958	1	US-08-456-837-6	Sequence 6, Appli
42	122.5	5.1	28958	1	US-08-457-342-6	Sequence 6, Appli
43	122.5	5.1	28958	1	US-08-457-646A-6	Sequence 6, Appli
44	122.5	5.1	28958	1	US-08-458-076A-6	Sequence 6, Appli
45	122.5	5.1	28958	1	US-08-764-233A-4	Sequence 4, Appli

ALIGNMENTS

RESULT 1

US-09-103-840A-2/c

; Sequence 2, Application US/09103840A

; Patent No. 6294328

; GENERAL INFORMATION:

; APPLICANT: FLEISCHMAN, Robert D.

; APPLICANT: WHITE, Owen R.

; APPLICANT: FRASER, Claire M.

; APPLICANT: VENTER, John C.

; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM

; TITLE OF INVENTION: TUBERCULOSIS

; FILE REFERENCE: 24366-2007.00

; CURRENT APPLICATION NUMBER: US/09/103, 840A

; CURRENT FILING DATE: 1998-06-24

; NUMBER OF SEQ ID NOS: 2

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 2

; LENGTH: 4403765

; TYPE: DNA

; ORGANISM: Mycobacterium tuberculosis

; FEATURE:

; OTHER INFORMATION: CDC 1551

; OTHER INFORMATION: "n" bases at various positions throughout the sequence

; OTHER INFORMATION: represent a, t, c or g

; US-09-103-840A-2

Alignment Scores:

Pred. No.: 0.0596

Score: 168.00

Percent Similarity: 37.67%

Best Local Similarity: 26.02%

Query Match: 6.97%

DB: 4

Length: 4403765

Matches: 96

Conservative: 43

Mismatches: 132

Indels: 98

Gaps: 14

US-09-825-414-7 (1-486) x US-09-103-840A-2 (1-4403765)

```

QY      145  ProAlaProLeuLysProLeuThrProLeuIleSer--GlyAlaLeuSerGlyAlaMetA 164
           |||||
Db 3936672 CCGGGCCCG-----CCGCTGCCCGCGCTGGTAGCTGAGAGAGCTGCCGGCGCTCCG 3936619
QY      164  spGlnValGlyThrLysMetMetAsparGAlaThrGlyAspLeuHisTyrLeuSerAlas 184
           |||||
Db 3936618 GTGCCCGCGCACCGCCTTGCCCGCGCTGCCAGCGGT----- 3936581
QY      184  erProAspArgLeuHisAspAlaMetAlaAlaSerValLysArgHisSerProSerLeuA 204
           :::::
Db 3936580 -----GCCGTTAGCGCCGCCACCGCGCGCCGACGCCCGGC----- 3936545
QY      204  IaArgGlnValLeuAspThrGlyValAlaValGlnThrTyrSerAlaArgAsnAlaValA 224
           |||||
Db 3936544 -----CGCCCCACCGGCACACC----- 3936527
QY      224  rgThrValLeuAlaProAlaLeuAlaSerArgProAlaValGlnGlyAlaValAspLeuG 244
           ||| ||||| ||| |||||
Db 3936526 -----GCTGCCGCTTGACCTCTTGCCCGCGCTTGCCGCCGGAGCTGGTGTGG 3936478
QY      244  lValSerMetAlaGlyGlyLeuAlaAlaAsnAlaGlyPheGlyAsnArg----- 260
           || :::: ||| ||| ||| ||| |||
Db 3936477 GGGCCGCT--GCCGCGCGCGCGCGCGCGCACCGCGGACTTGCCGCTCCGCGCTGGCCG 3936421
QY      261  --LeuLeuSerValGlnSerArgAspHisGlnArgGlyGlyAlaLeuValLeuGlyLeuL 280
           ||| :::: :::: ||| |||||
Db 3936420 CCTTGCGCGCGCGCGCACCGACCGATACCGTGCGCGCGCGCGCCACCACT----- 3936371
QY      280  ysAspLysGluProLysAlaGlnLeuSerGluGluAsnAspTrpLeuGluAlaTyrLysA 300
Db 3936371 ----- 3936371
QY      300  IaIleLysSerAlaSerTyrSerGlyAlaAlaLeuAsnAlaGlyLysArgMetAlaGlyL 320
           ::||| ||| ::||| ||||| :::: |||||
Db 3936370 -----GCCGCTCTGCCCGCGCGGTGCCGCGCTTGCCCGCGCGCGAGCCGGTGCCGCC 3936319
QY      320  euProLeuAspMetAlaThrAspAlaMetGly--AlaValArgSerLeuValSerAlas 339
           :::: :::: ||| ||| :::: |||
Db 3936318 CCACCGTTGCTGCCAGCGGTGGCGCGCGCGCGCGAGCGCGCGCTTGCCGCGCTGCCGCCG 3936259
QY      339  erSerLeuThrGlnAsnGlyLeuAlaLeuAlaGly--GlyPheAlaGlyValGlyLysL 358
           ::::: ||||| ||::: ||| ||| ||| |||
Db 3936258 GTACCGTTGGCGCGCGAGGTACCGCGGTGCTGCCCGCGCGCACACCGCGACCGCGGT 3936199
QY      358  euGlnGluMetAlaThrLysAsnIleThrAspProAlaThrLysAlaAlaValSerGlnL 378
           ||::: ||| ||| ||||| ::::: |||
Db 3936198 CCGCCCTTCCGCCAGT-----CCACCGTTTCCGCCGCTATTCCGCTA 3936154
QY      378  euThrAsnLeuAlaGlySerAlaAlaValPheAlaGlyTrpThrThrAlaAlaLeuThrT 398
           ||| ::::: ||||| ||| ||| ||||| |||
Db 3936153 CCACCGCGGTCTCCGCCAGCGCGCGCGATGGCGCGCTTGCCCGCGGTGCGCGGTACT 3936094
QY      398  hrAspProAlaValLysLysAlaGluSerPheIleGlnAspThrValLysSerThr---- 416
           :: ||||| ::::: ||| ||| |||||
Db 3936093 CCGGTCCCGCGCGCGAGCGCGTGCATGCCGCTACCGCGACGACCGCGGTTCACCGTTG 3936034
QY      417  -----AlaSerSerThrThrGlyTyrValA 425
           |||||
Db 3936033 CCGCGCGCGCGCGGTTCGCCCATGCCCGCGCGCTGCCAGCGCGGTGCGCGCGAGATG 3935974
QY      425  IaAspGlnThrValLysLeuAlaLysThrValLysAspMetGlyGlyAlaIle---- 443
           ||||| ||| :::: ||| :::: ||| ::::
Db 3935973 CCGACGCGCACCGCGCATCTCCGCGCTGCCCGCATTGCCAGCATACCGTTGGCG 3935914
QY      444  -----ThrHisThrGlyAlaSerLeuArgAsnThrValAsnLeuArgGlnArgProA 462
           :::: ||||| :::: ||| ||::: |||
Db 3935913 CCGGAGTGGCACCGCGCGCGC-----CGCGCGCGCGTCTCCGCCCTTCCGCCCGCTGCCG 3935857
QY      462  IaArgGluAlaAspIleGluGly-----GlyThrAlaAlaSerP 476
           ||||| ||| :::: ||| ||| |||||
Db 3935856 CCGCGCGCGCGCGGTGGCGCGGTGCTGCTGTGACCGCTCTCCACCGTCCGCCCT 3935797
QY      476  roSerGluIleProPheArgPro 483

```

```

Db 3935796 TTGCCGCATTTGCCGCCGCC 3935774
::: ||| |||||
RESULT 2
US-09-103-840A-1/c
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: Patentln Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
US-09-103-840A-1

```

[illegible]

US-09-825-414-7 (1-486) x US-09-103-840A-1 (1-4411529)

QY	145	ProAlaProLeuLysProLeuThrProLeuIleSer--GlyAlaLeuSerGlyAlaMet	164
Db	3943125	CCGGCCGC-----CCGCTGCCCGCTGTGGTAGCTGAGAGAGCTGCCGCCGCCGCCG	3943072
QY	164	spGlnValGlyThrLysMetMetAspArgAlaThrGlyAspLeuHisTyrLeuSerAla	184
Db	3943071	GTGCCCGCCGACCCGCTTGCGCGCGCTTGCCACGCGT-----	3943034
QY	184	erProAspArgLeuHisAspAlaMetAlaAlaSerValLysArgHisSerProSerLeu	204
Db	3943033	-----GCCGTTAGCGCCGCCACCGCGCCGCCACAGCCCGGC----	3942998
QY	204	laArgGlnValLeuAspThrGlyValAlaValGlnThrTyrSerAlaArgAsnAlaVal	224
Db	3942997	-----CGCCCCACCCGGCACACC-----	3942980
QY	224	rgThrValLeuAlaProAlaLeuAlaSerArgProAlaValGlnGlyAlaValAspLeu	244
Db	3942979	-----GCTGCCGCTTGACTCTCTTGCCCCCGCTGGCCCGGAGCTGGTGTG	3942931
QY	244	lyValSerMetAlaGlyGlyLeuAlaAlaAsnAlaGlyPheGlyAsnArg-----	260
Db	3942930	GGCCCGCT---GCCCGCGCGCGCCGCCACCGCGCACTTGCGCGCGCGCTGCCG	3942874
QY	261	--LeuLeuSerValGlnSerArgAspHisGlnArgGlyAlaLeuValLeuGlyLeu	280
Db	3942873	CCCTTGCCGCGCCGACGACGACAGATACCGTGCAGCGCGCGCCACCACT-----	3942824
QY	280	ysAspLysGluProLysAlaGlnLeuSerGluGluAsnAspTrpLeuGluAlaTyrLys	300
		-----	-----
Db	3942824	-----	3942824
QY	300	laileLysSerAlaSerTyrSerGlyAlaAlaLeuAsnAlaGlyLysArgMetAlaGlyL	320
Db	3942823	-----GCCGCTCTGCGCGCGGCTGCCGCGCTTGCGCGCGCGGAGCCGCGTCCG	3942772
QY	320	euProLeuAspMetAlaThrAspAlaMetGly--AlaValArgSerLeuValSerAla	339


```

      ::::
Db 3942771 CCACCGTTGCTGCCAGCCCGTGGCGGCCCCCGGCGACGCGCCCGCTTGCCGCGCTCCGCGCG
QY 339 erSerleuThrGlnAsnGlyLeuAlaLeuAlaGly--GlyPheAlaGlyValGlyLysL 358
      :::::
Db 3942711 GTACCGTTGGCGCCGACGCGTACCGCGTGTGTCGCCGCGCACACCAGCGGACCGGC---- 3942656
QY 358 euGlnGluMetAlaThrLysAsnIleThrAspProAlaThrLysAlaAlaValSerGlnL 378
      :::
Db 3942655 -----GTTCCGCGCTTTCGCCCAATCCACCGTTTCCGCGCTATTTCGCTA 3942607
QY 378 euThrAsnLeuAlaGlySerAlaAlaValPheAlaGlyTrpThrThrAlaAlaLeuThrT 398
      ||| ::::: ||||| |||
Db 3942606 CCACCGCGCTCTCCGCGCACGCGCGGATGGCGCGCTTGCCGCGCTGCGCGGTACTT 3942547
QY 398 hrAspProAlaValLysLysAlaGlySerPheIleGlnAspThrValLysSerThr---- 416
      :: ||||| ::::: |||
Db 3942546 CCGGTCCCGCGCGCGACGCGCTGCATGCCGTACCGCGCACGACCGCGGTTCACCGCTT 3942487
QY 417 -----AlaSerSerThrThrGlyTyValA 425
      |||||:::
Db 3942486 CCGCGCGCGCGCGGTTCCGCCATGCCCGCGCTGCCACGCGCGGTGCGCGGAGATG 3942427
QY 425 laAspGlnThrValLysLeuAlaLysThrValLysAspMetGlyGlyAlaIle---- 443
      |||| ||| ::| |||
Db 3942426 CCGACGGGCACACGCGCGCGCATCTCCGCGCGCGCATGCGACGATACCGTTGGCG 3942367
QY 444 -----ThrIsthrGlyAlaSerLeuArgAsnThrValAsnAsnLeuArgGlnArgProA 462
      ::| |||||::: ||| ||::: |||
Db 3942366 CCGCGAGTGGACCGCGCGCGCC--CGCGCGCGCTCTCGCGCTTCCGCGCGTGGCG 3942310
QY 462 laArgGluAlaAspIleGluGly-----GlyThrAlaAlaSerP 476
      |||| ||| ::| |||
Db 3942309 CCGCGCGCGCGCTGTGGCGGGGTGCTGTGTGACGCGCTGCTCCACCGTCCGCGCCT 3942250
QY 476 roSerGluIleProPheArgPro 483
      ::| ||| |||||
Db 3942249 TTGCCGCCATTGCGCGCGCGCC 3942227

RESULT 3
US-08-119-125A-2
; Sequence 2, Application US/08119125A
; Patent No. 5610011
; GENERAL INFORMATION:
; APPLICANT: SMITH, Hilda Elizabeth
; APPLICANT: VECHT, Uri
; TITLE OF INVENTION: DNA Sequences which code for Virulence
; TITLE OF INVENTION: Characteristics of Streptococcus suis and parts thereof, polyP
; TITLE OF INVENTION: antibodies derived therefrom and the use thereof for the diagn
; TITLE OF INVENTION: protection against infection by S. suis in mammals, including
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Centraal Diergeneeskundig Instituut
; STREET: Edelhertweg 15
; CITY: PH Leystad
; STATE:
; COUNTRY: The Netherlands
; ZIP: NL-8219
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.44 MB storage
; COMPUTER: IBM compatible
; OPERATING SYSTEM: MS-DOS v.6.0
; SOFTWARE: Wordperfect v. 6.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/119,125A
; FILING DATE: 20-SEP-1993
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/NL92/00054
; FILING DATE: 19-MAR-1992
; APPLICATION NUMBER: NL 9100510
; FILING DATE: 21-MAR-1991
; ATTORNEY/AGENT INFORMATION:
```

```

; NAME: Handal, Anthony H.
; REGISTRATION NUMBER: 26275
; REFERENCE/DOCKET NUMBER: SMITHHE119125
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (203) 838-8589
; TELEFAX: (203) 838-8794
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6744 base pairs
; TYPE: Nucleic acid with corresponding amino acids
; STRANDEDNESS: single stranded
; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA
; ORIGINAL SOURCE:
; ORGANISM: Streptococcus suis type II (pathogenic)
; FEATURE:
; OTHER INFORMATION: Extracellular factor related protein (EF*) gene
; FEATURE:
; NAME/KEY: promoter -35 region
; LOCATION: bp 66 to 71
; FEATURE:
; NAME/KEY: promoter -10 region
; LOCATION: bp 89 to 94
; FEATURE:
; NAME/KEY: promoter -35 region
; LOCATION: bp 153 to 158
; FEATURE:
; NAME/KEY: promoter -10 region
; LOCATION: bp 176 to 181
; FEATURE:
; NAME/KEY: ribosome binding site
; LOCATION: bp 350 to 356
; FEATURE:
; NAME/KEY: signal peptide
; LOCATION: bp 361 to 498
; FEATURE:
; NAME/KEY: start of repetitive units R1-R11
; LOCATION: bp 2869, 3097, 3292, 3520, 4087, 4381, 4609, 4837,
; LOCATION: 5065, 5293, 5521:
; FEATURE:
; NAME/KEY: start of repetitive Asn-Pro-Asn-Leu sequences
; LOCATION: bp 2932, 3160, 3355, 3583, 4150, 4444, 4672, 4900,
; LOCATION: 5128, 5356, 5584:
; FEATURE:
; NAME/KEY: dyad symmetry regions
; LOCATION: from bp 6554 to 6566 and from bp 6571 to 6583
; FEATURE:
; NAME/KEY: dyad symmetry regions
; LOCATION: from bp 6611 to 6625 and from bp 6631 to 6644
; US-08-119-125A-2

Alignment Scores:
Pred. No.: 0.000199 length: 6744
Score: 147.50 Matches: 105
Percent Similarity: 36.76% Conservative: 81
Best Local Similarity: 20.75% Mismatches: 219
Query Match: 6.12% Indels: 101
DB: 1 Gaps: 19

US-09-825-414-7 (1-486) x US-08-119-125A-2 (1-6744)
QY 14 AlaThrAspSerPheArgThrAlaSerAspAlaSerLeuAlaSerSerSerValArgSer 33
      ||| ||::: ||| ||| ||||| :::::|:::
Db 4024 GCTAGAGTGCAGTGTGATTTGGCTAAGCATTAAGAATTAGCTAAGAGAACAATCCGACACA 4083
QY 34 ValSerSerAspGlnArgGluIleAsnAlaIleAlaAspTyrLeuThrAspHisVal 53
      ::| ::| ::| ::| ::| ::| ::|
Db 4084 GAAGAGAAGAAGCTACTAAATATGTAAGAAACTTGCAGAA-----GATACGCGC 4134
QY 54 PheAlaAlaHisLysLeuProProAlaAspSerAlaAspGlyGlnAlaValAspVal 73
      ||| ||| ||| ||| ||||| ::|
Db 4135 AAAGCTATCGAGGACAATCCAACTTGTCAAGATGAAGATGAAGCAAGCGGAATTAAAG 4194
```

QY 74 Hisasn-----AlagInileThralaleuIlegluThrArgAlaSerArgLeuHis 90
Db 4195 CTAACGTACGCTGTGGCAAAAACCTTTAGCAACCATTCGTGACAAATGCAGATTAAGCGTACG 4254
QY 91 PhegluGlygluThrProAlaThrIleAlaaspthrPheAlaLysAlagluysleuasp 110
Db 4255 CAAGAAGCAGAAAAAGCTCAAGCCCTAGCAGAT--CTTGAAAAAGCTAAAGAAACACAG 4311
QY 111 ArgLeuAlaThrThrThrSer-----GlyAlaLeu 120
Db 4312 AAAATTCAGATAAAGCTGCATTCGATAGTTGACTPACTTGTGAAGATGGTGAGCTT 4371
QY 121 ArgAlaThrProPheAlaMetAlaSerLeuGlnTyrMetGlnProAlaIleAsnLys 140
Db 4372 GAAGCTACT-----AAACAAGATGCTAAGAACAG 4401
QY 141 -----GlyAspTyrLeuPro 145
Db 4402 ATTGCTAAAGATGCAGCCGCTGCTAAAGAAGCAATTGCCAAGCAATCCAAACTTGACAGAC 4461
QY 146 AlaProLeuLysProLeuThrProLeuIleSerGlyAlaLeuSerGlyAlaMetaspGln 165
Db 4462 GCAGAGAAGAAAACTTCACCGATGCGGTAGATGCAGAAAGTAGCCAAAGCTAACGACGCA 4521
QY 166 ValGlyThrLysMetMetaspArgAlaThrGlyAspLeuHisTyrLeuSerAlaSerPro 185
Db 4522 ATTTCAGCT-----GCAACCAGC-----CCAGCAGATGTT 4551
QY 186 AspArgLeuHisaspAlaMetAlaAlaSerValLysArgHisSerProSerLeuAlaArg 205
Db 4552 CAAAAAGAAGAGATGCGAGTGTTCAGCCATTCAGAAAGATGTTCTTGACGCGCTTAA 4611
QY 206 GlnValLeuAspThrGlyValAlaValaGlnThrTyrSerAlaArgAsnAlaValaArgThr 225
Db 4612 CAAGATGCTAAGATAAGATTGCTTAAAGATGCAGCCGCTGCTTAAAGAAGCAATGGCTCC 4671
QY 226 ValLeuAlaProAlaLeuAlaSerArgProAlaValGlnGlyAlaValaAspLeuGlyVal 245
Db 4672 AATCCAAACTTGACAGACGCAGAGAAGAAAAACCTTCACCGATGCGGTAGATGCAGAAAGTA 4731
QY 246 SerMetAlaGlyGlyLeuAlaAlaAsnAlaGlyPheGlyAsnArgLeuLeuSerValGln 265
Db 4732 GCCAAAGCTAACGACGCAATTCAGCTGCA-----ACCAGCCACGAGATGTTCAA 4782
QY 266 SerArgaspHisGlnArgGlyGlyAlaLeu-----ValLeuGlyLeuLysAspLys 282
Db 4783 AAAGAAGAGGATGCAGGTGTTGCAGCCATTCGACAGAGATGTTCTTGACGCGAGCTTAAACAA 4842
QY 283 GluProLysAlaGlnLeuSerGluGluAsnAspTyrPheGluAlaTyrLysAlaIleLys 302
Db 4843 GATGCTAAGATAAGATTGCTTAAAGATCCGAC-----GCTGCTAAG 4884
QY 303 SerAla-----SerTyrSerGlyAlaAlaLeuAsnAlaGlyLysArgMet 317
Db 4885 TCAGCCATTGACGCGAAATCCAAACTTGACAGATGCAGAGAAGAAATCAGCTAAGAAAGCA 4944
QY 318 AlaGlyLeuProLeuaspMetAlaThrAspAlaMetGlyAlaValaArgSerLeuValSer 337
Db 4945 GTTGATGCTGATGCTAAAGCTGCGACAGATGCATTCGTTCAACAAGTCCAGTCGAA 5004
QY 338 AlaSerSerLeuThrGlnAsnGlyLeuAlaLeuAlaGlyGlyPheAlaGlyValGlyLys 357
Db 5005 GCGCAATCGGACAGAGACAA-----GCGGTAGCTTCA 5037
QY 358 LeuGlnGluMetAlaThrLysAsnIleThrAspProAlaThrLysAlaAlaValaSerGln 377
Db 5038 ATC-----GCCCAAGATGTTCTTGACGCGAGCAACAAGATGCTAAGAACAAAG 5085
QY 378 LeuThrAsnLeuAlaGlySerAlaAlaValaPheAlaGlyTyrPThrThraIaAlaLeuThr 397
Db 5086 ATTGCCAAAGAAGTTGCCGACGCT-----AAAGAAGCAATTGAT 5124
QY 398 ThrAspProAlaValLysLysAlaGlu--SerPheIleGlnAspThrValLysSerThr 416

Db 5125 GCCAATCCGAACCTTATTCAGATGCAGAGAAGAAAGCTTCTAAGAAAGCGGTAGATGCAGAT 5184
QY 417 AlaSerSerThrThrGlyTyrValAlaAspGlnThrValLysLeu--AlaLysThrVal 435
Db 5185 GCTAAAGCTACGACAGATGCATTCATTCGTTCAACAAGTCCAGTCCGAAGCGCATCGGCA 5244
QY 436 LysAspMetGlyGlyGluAlaIleThrHisThrGlyAlaSerLeuArgAsnThrValasn 455
Db 5245 GAGGACAAAGGC-----GTAGGTTCAATCGCCCAAGATGTTCTTGAC 5286
QY 455 AsnLeuArgGln-----ArgProAlaArgGluAlaAspIleGluGlyGly 471
Db 5287 GCAGCGAAACAAGATGCTAAGATAAGATTGCTTAAAGAATCCGACGCTGCTAAGTCAGCC 5346
QY 472 ThrAlaAlaSerProSer 477
Db 5347 ATTGACGCGAATCCAAAC 5364
RESULT 4
US-09-453-702B-72
; Sequence 72, Application US/09453702B
; Patent No. 6365723
; GENERAL INFORMATION:
; APPLICANT: Blattner, Frederick R.
; Burland, Valerie
; Perna, Nicole T.
; Plunkett, Guy
; Welch, Rod
TITLE OF INVENTION: No. 6365723e1 Sequences of E. coli O157
NUMBER OF SEQUENCES: 265
CORRESPONDENCE ADDRESS:
ADDRESSEE: Quarles & Brady
STREET: 1 South Pinckney Street
CITY: Madison
STATE: WI
COUNTRY: US
ZIP: 53701-2113
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.44mb storage
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Word Perfect 8.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/453,702B
FILING DATE: 03-Dec-1999
CLASSIFICATION: <unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/110,955
FILING DATE: 04-DEC-1998
ATTORNEY/AGENT INFORMATION:
NAME: Seay, Nicholas J.
REGISTRATION NUMBER: 27386
REFERENCE/DOCKET NUMBER: 960296.95017
TELECOMMUNICATION INFORMATION:
TELEPHONE: (608) 251-5000
TELEFAX: (608) 251-9166
; INFORMATION FOR SEQ ID NO: 72:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 46819
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; SEQUENCE DESCRIPTION: SEQ ID NO: 72:
US-09-453-702B-72
Alignment Scores:
Pred. No.: 0.0118 Length: 46819
Score: 143.50 Matches: 107
Percent Similarity: 33.40% Conservative: 52
Best Local Similarity: 22.48% Mismatches: 232
Query Match: 5.95% Indels: 85


```
DB:          4          Gaps:          15
US-09-825-414-7 (1-486) x US-09-453-702B-72 (1-46819)
QY      19 ArgThrAlaSerAspAlaSerLeuAlaSerSerSerValArgSerValSerSerAspGln 38
      ||| ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      32015 CGTAACGCGTCCGCGGTGGGCACAGAACACGCGCCGAGAGAAGTCAGCCAGCATGCC 32074
QY      39 GlnArgGluIleAsnAlaIleAlaAspTyrLeuThrAspHisValPheAlaAlaHisLys 58
      ||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      32075 AGCACATCAGCCCGGTGAGGGCGCAACCCATGCCACTGAT-----GCTGCGGACTCA 32125
QY      59 LeuProProAlaAspSerAlaAspGlyGlnAlaAlaValAspValHisAsnAlaGlnIle 78
      ||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      32126 GCACGCGCAGCCAGCACGTCAGCCGCGACAGCGCGCTCGTCGGCTCAGTCAGCGTCTTCC 32185
QY      79 ThrAlaLeuIleGluThrArgAlaSerArgLeuHisPheGluGlyGluThrProAlaThr 98
      : : : ||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      32186 AGCGCA-----GGAACGGCATCAACGAAGGCTTACT 32215
QY      99 IleAlaAspThrPheAlaLysAlaGluLysLeuAspArgLeuAlaThrThrSerGly 118
      ||| ||| ||| : : : : : : : : : : : : : : : : : : : : :
Db      32216 GAAGCATCAAAAAGTCTGCCGCTGCAGAGTCTCAAAAAGCGCGCGCTACCAGTGCC 32275
QY      119 AlaLeuArgAlaThrProPheAlaMetAlaSerLeuLeuGlnTyrMetGlnProAlaIle 138
      ||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      32276 GGTGCGGCGAAGCGTCAGAAACGAATGCGGCAGTGCACAA-----CAATCAGCCGCC 32329
QY      139 AsnLysGlyAspTyrPleuProAlaProLeuLysProLeuThrProLeuIleSerGlyAla 158
      ||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      32330 ACTTCTGCA-----TCCACCGCGAGCACCAAGCGTCAAGAAGCT 32368
QY      159 LeuSerGlyAlaMetAspGlnValGlyThrLysMetMetAspArgAlaThrGlyAspLeu 178
      ||| ||| ||| : : : ||| : : : : : : : : : : : : : : : : : : : :
Db      32369 GCCTCCTCAGCCAGGGATGCGTCGGCTTCAAAAAGCGCGCAAAATCATCA---GAAACG 32425
QY      179 HisTyrLeuSerAlaSerProAspArgLeuHisAspAlaMetAlaAla-----SerVal 196
      ||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      32426 AGCGAGCCTCGAGCGCCAGTAGTGACAGCCTTCCTCGGCAACGCGCGCAATTCGCGG 32485
QY      197 Lys-----ArgHisSerProSerLeuAlaArgGlnVal 207
      ||| : : : ||| : : : ||| : : : |||
Db      32486 AAGCGGCGCAAAACGTTCTGAGACAAACGCTAAGTCTCTGAAACGGCAGCAACAGAGT 32545
QY      208 LeuAspThrGlyValAlaValAlaGlnThrTyrSerAla-----ArgAsnAlaValArgThr 225
      : : : ||| : : : ||| : : : ||| : : : |||
Db      32546 GCCTCCGCGACGACGAGGCTCAAAAACAGCGGCTGCATTTATCTGCCAGTCCGCGTCAACA 32605
QY      226 ValLeuAlaProAlaLeuAlaSerArgProAlaValAlaGlnGlyAlaValAspLeuGlyVal 245
      ||| ||||| ||| : : : : : : : : : : : : : : : : : : : :
Db      32606 AGTCCGGGCGAGGCTTCAGCCAGTGCACCGCGCGGAAATCGGCAGAAAGTGCCGCA 32665
QY      246 SerMetAlaGlyGlyLeuAlaAlaAsnAlaGlyPheGlyAsnArgLeuLeuSerValGln 265
      ||| ||| ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      32666 TCGTCTGCTTCAACAGCCACAAAGAGGCTGGCGAAAGCCACTGAACAGCGCAGCGCA 32725
QY      266 SerArgAspHisGlnArgGlyGlyAlaLeuValLeuGlyLeuLysAspLysGluProLys 285
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      32726 GCGAGTTCCTGCTTCGCGACGAGACATCCGAACGAACGCGAAAGCGTTGGAACACGAGC 32785
QY      286 AlaGlnLeuSerGluGluAsnAspTyrPleuGluAlaTyrLysAlaIleLysSerAlaSer 305
      ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : |||
Db      32786 GCAGAATCCTCAAAAACGGCTGCCGCATCCTCAGCCAGTTCGGCGCGCTCATCGGCATCA 32845
QY      306 TyrSerGlyAlaAla-----LeuAsnAlaGlyLysArgMetAla 318
      : : : ||| : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      32846 TCTGCGTCTGTTCAAAAAGATGAGGCGACCAACAAGCGTCAGCAGCGAAGACAGCGGCC 32905
QY      319 GlyLeuProLeuAspMetAlaThrAspAlaMetGlyAlaValArgSerLeuValSerAla 338
      ||||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : |||
Db      32906 ACGACGGCATCCAGGAAGCGCAGACAGAGGCACTGTGAGCAGCGCAGCTCAGAGC 32965
QY      339 SerSerLeuThrGlnAsnGlyLeuAlaLeuAlaGlyGlyPheAlaGlyValGlyLysLeu 358
```

```
Db      32966 AAAAGTACGGCGGATCTGCAGCAACGCGCGCT----- 32998
QY      359 GlnGluMetAlaThrLysAsnIleThrAspProAla----- 370
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db      32999 ---GAGACAGCGCGCAAAACGGGCGAGAGGATATTGCATCCGCCGTGAGGATGCC 33055
QY      371 -----ThrLysAlaAlaValSerGlnLeuThrAsnLeuAlaGlySerAlaAlaValPhe 388
      ||||| ||||| : : : ||||| : : : ||||| : : : ||||| : : : ||||| : : : |||||
Db      33056 AGCAGCAGCAAAAAAGGGATAGTACAGCTCAGCAGTGCAGCTAACAGC----- 33103
QY      389 AlaGlyTyrThrThrAlaAlaLeuThrThrAspPro---AlaValLysLysAlaGluSer 407
      ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : |||
Db      33104 -----ACTTCCGAGTCACTGGCGGCAACGCCAAAAGCCGTTAAGCGCGCTATGAG 33154
QY      408 Phe-----IleGlnAspThrValLys----- 414
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      33155 CTGGCTAACGGGAATACACCCGACAGGATGCACACAGACAGCAAGAAGGATAGTTCAG 33214
QY      415 ---SerThrAlaSerSerThrThrGlyTyrValAlaAspGlnThrValLysLeuAlaLys 433
      ||| ||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      33215 CTTAGCAACGCGACCAACAGCACATCTGAAATGCTGGCGGCAACGCCAAAGTGGTAAAG 33274
QY      434 ThrValLysAspMetGlyGlyGluAlaIleThrHisThrGlyAlaSer 449
Db      33275 GCAGCCTATGACCTTGCCTAACGGGAATATACTGCTCAGGACGCTACG 33322

RESULT 5
US-09-103-840A-1
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
US-09-103-840A-1

Alignment Scores:
Pred. No.:      60.7      Length:      4411529
Score:          138.50    Matches:      122
Percent Similarity: 33.91% Conservative: 53
Best Local Similarity: 23.64% Mismatches: 193
Query Match:      5.74%   Indels:      148
DB:              4      Gaps:      21

US-09-825-414-7 (1-486) x US-09-103-840A-1 (1-4411529)
QY      20 ThrAlaSerAspAlaSerLeuAlaSerSerSerValArgSerValSer----- 35
      ||||| ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : |||
Db      1632232 ACCGCCCTGCCCGCGCGCGCGGGGTTACCGCGGTTAACCGCGTGACCGCGCGGTTAACCATC 1632291
QY      36 SerAspGlnGlnArgGluIleAsnAlaIleAlaAspTyrLeuThrAspHisValPheAla 55
      : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      1632292 GCCGAAGCGGAAGTGCCTGGTGGCGCGGTTGCCCGCTCACCGCGGAGCGCGCGCGCC 1632351
QY      56 AlaHisLysLeuProProAlaAspSerAlaAspGlyGlnAlaAlaValAspValHisAsn 75
      ||||| ||| ||||| ||| : : : ||| : : : ||| : : : ||| : : : ||| : : : |||
Db      1632352 CCCTTGGCCGCGCGGACCGCGGACACACCCCTGATTCGTTCTGGCCAAAGAGTTCCCGGC 1632411
QY      76 -----AlaGlnIleThrAlaLeuIleGluThrArgAlaSerArg 88
```

Db	1632412	CAAACGCGCGGGCCGCTTGGCCGCCGTATACGCGCTTG-----CGCGCC-----	1632456
QY	89	LeuHisPheGluGlyIuThrProAlaThrIleAlaAspThrPheAlaLysAlaGluLys	108
Db	1632457	-----GGGCGCGCGGTGGCGCGCGGTGCSCGCGCCTTGGCGCGCG--	1632495
QY	109	LeuAspArgLeuAlaThrThrThrSerGlyAlaLeuArgAlaThrProPheAlaMetAla	128
Db	1632496	-----CGTGGTGGCGGTGGCGCGCGGTGGCGCGCGTTGGCGCGCGCCACCAGCTCCGCGC	1632549
QY	129	SerLeuLeuGlnTyTMetClnProAlaIleasnLysGlyAspTrpLeuProAlaProLeu	148
Db	1632550	GTCGCCCCCGAAGTCTCCGCCCCGCGC-----GCCGCGCGCGCC	1632588
QY	149	LysProLeuThrProLeuIleSerGlyAlaLeuSerGlyAlaMetaspGlnValGlyThr	168
Db	1632589	CGCCCCCGCAGCCCCCGCGCTTCTGGCGCGCTCGTGCGGATTGCGCGCGGTGGTGG---	1632645
QY	169	LysMetMetaspArgAlaThrGlyaspLeuHisTyTrLeuSerAlaSerProaspArgLeu	188
Db	1632646	-----CGAGGAACCGGC-----	1632657
QY	189	HisaspAlaMetAlaAlaSerValLysArgHisSerProSerLeuAlaArgGlnValLeu	208
Db	1632657	-----	1632657
QY	209	AspThrGlyValAlaValGlnThrTyTrSerAlaArgaspAlaValArgThrValLeuAla	228
Db	1632658	GACACCGGCGCATGCGCGTCCCGCCTTGTGCGCGC-----GGC	1632693
QY	229	ProAlaLeuAlaSerArgProAlaValGlnGly-----Ala	240
Db	1632694	CCC GCCATTACCAACAAGCCCCGCTTGC CGCCCTTGCCGCGCGCGCGCGCGCGCC	1632753
QY	241	ValaspLeuGlyValSer-----MetAlaGlyGlyLeuAlaAlaAsnAlaGlyPheGly	258
Db	1632754	GCGCACGGGTGGCGCTTCCGCGCGCTTGCCGCGCGGTGCGCGCGCGCTGCGCGGT	1632813
QY	259	AsnaArgLeuLeuSerValGlnSerArgaspHisGlnArgGlyGlyAlaLeuValLeuGly	278
Db	1632814	GCGCGCGCGGCGCACCGCTGTGCACCCGCGGTGATCC--GCGCGCGCGATCCACACAGC	1632870
QY	279	LeuLysaspLysGluProLysAlaGlnLeuSerGlnGluAsnaspTrpLeuGluAlaTyr	298
Db	1632871	ACCACCGATGCGCGCGCTAACCGCGCTTGCGCGCGCTTGCCACCACTCATCGCGCGCTT	1632930
QY	299	-----LysAlaIleLysSerAlaSerTyTrSerGlyAlaAlaLeu	311
Db	1632931	ATCGAAGCTGCCCTTGGCACCGCTTGCGGCATCACCGCCATGCGCGCGCGCGCGCTT	1632990
QY	312	AsnaIagLy-----LysArgMetaAla	318
Db	1632991	TCCGCGCGCGCGCGCGCGCACCCATGCTGCCGCTCCTGTGGGTGGCTGCAAGCGCCTTACC	1633050
QY	319	GlyLeuProLeuaspMetaIaThraspAlaMetGlyAlaValArgSerLeuValSerAla	338
Db	1633051	GCGTTGCCACCGCGGTCCACCGCACCGCGCGCTCCACCGTTGCCCGCTTGCGCGCGCT	1633110
QY	339	SerSerLeuThrGlnAsnGlyLeuAlaLeuAlaGlyGlyPheAlaGlyValGlyLysLeu	358
Db	1633111	GGTGCCATCCGCGCTGCCCGCCAGCGCGTTAAGGCGCGGTGGCGCGCGT-----	1633158
QY	359	GlnGluMetaIaThrLysAsnIleThraspProAlaThrLysAlaAlaValSerGlnLeu	378
Db	1633159	-----GGCGCGCTTGCCGCGCGCTT	1633185
QY	379	ThrasnLeuAlaGlySerAlaAlaValPheAla--GlyTyTrThrThraAlaLeuthr	397
Db	1633186	ACCGCGCGCGCGCGCGCAGACCGCGCTGCGCTTGCGCTCGCGCGCGCGCTTACC	1633245
QY	398	ThraspProAlaValLysLysAlaGluSerPhe-----IleGlnAspThr	412

```

Db 1633246 GCCAGCGCCGCCAGCTCCGCCGCCACCCGCTTAGGGTCGCCGCCGACAGAGCGGGGGACC 1633305
QY 413 VallysSerThrAlaSerSerThrThrgly---TyrValAlaAspGlnThrValLysLeu 431
Db 1633306 GGGGGCGCGCTGTGCCGCCGCCGACCTCCGGCGCCGCCCATTTGCCGACACGCCCGCGGCC 1633365
QY 432 AlaLysThrValLysAspMetGlyGlyGlnAlaIleThrHisThrglyAlaSerLeuArg 451
Db 1633366 GCC-----GGCCCCGCGCTTACC---GCCGGCTTTGCCGCCCGCA 1633401
QY 452 AsnThrValAsnAsnLeuArgGlnArgProAla----- 462
Db 1633402 TGAGAAGTGGCGCGCTTGCCGCCGCCCGCGCTTGCCGCCCGCTGTGGGGCTGGCCCC 1633461
QY 463 ---ArgGluAlaAspIleGluGlyGlyThrAlaAlaSerProSer 477
Db 1633462 GCGCGCGCGCTGGCGCACCGATCGTGAGCGCGCGCTCCGCGGTGCTCC 1633509
RESULT 6
US-08-961-527-71
; Sequense 71, Application US/08961527
; Patent No. 6420135
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 391
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,527
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 71:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 32768 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-08-961-527-71
Alignment Scores:
Pred. No.: 0.027 Length: 32768
Score: 137.50 Matches: 103
Percent Similarity: 39.24% Conservative: 83
Best Local Similarity: 21.73% Mismatches: 243
Query Match: 5.70% Indels: 47
DB: 4 Gaps: 13
US-09-825-414-7 (1-486) x US-08-961-527-71 (1-32768)
QY 1 MethiSileasnArgArgValGlnGlnProProValThrAlaThrAspSerPheArgThr 20
:::|||||:::|||||:::|||||:::|||||:::|||||
Db 4488 GTGCATCTGAATCG-----CATCAACCACT-GCGTCAGCCTCAGCAAGTACTAGCGCC 4540

```


QY 21 AlAserAspAlaSerLeuAlaSerSerValArgSerValSerSerAspGlnArg 40
Db 4541 TCAGCCTCAGCATCAACGAGTGCCTCCGCTTCAGCAAGTACTAGTCATCAGCTTCAGCA 4600
QY 41 GluileAsnAlaIleAlaAspTyrLeuThrAspHisValPheAlaAlaHisLysLeuPro 60
Db 4601 AGTACTAGCGCCTCAGCCTCAGCGTGCAGCAAGCGCCTCAGCTTCAGCA----- 4648
QY 61 ProAlaAspSerAlaAspGlyGlnAlaAlaValAspValHisAsnAlaGlnIleThrAla 80
Db 4649 ---AGTACCAGTGCCTCAGCGCTCAGCGTGCAGCAAGTGCCTTCAGCAAGTACTCA 4705
QY 81 LeuileGluThrArgAlaSerArgLeuHisPheGluGlyGluThrProAlaThrIle--- 99
Db 4706 GCGTCTGAATCAGCATCAACAAGTGCCTCGGCTTCAGCATCAACGAGTGCATCAGCTTCA 4765
QY 100 AlaAspThrPheAlaLysAlaGluLysLeuAspArgLeuAlaThrThrThrSerGlyAla 119
Db 4766 GCATCAACAAGTGCCTTCAGCTTCA-----GCAAGTACCAGTGCCTCGGCT 4810
QY 120 LeuArgAlaThrProPheAlaMetAlaSerLeuLeuGlnTyrMetGlnProAlaIleAsn 139
Db 4811 TCAGCATCAACGAGTGCCTCAGTCTCAGCGTCAACCAAGTGCCTTGAAATCCGCAACA 4870
QY 140 LysGlyAspTyrLeuProAlaProLeuLysProLeuThrProLeuIleSerGlyAlaLeu 159
Db 4871 AGT-----GCCTTCGCGTTCAGCAAGCAACAGTGCCTTCGCGTTCAGCGTCA 4915
QY 160 SerGlyAlaMetAspGlnValGlyThrLysMetMetAspArgAlaThrGlyAspLeuHis 179
Db 4916 ACGAGTGCCTGAGTCAGCATCAACGAGTGCCTCAGCGCTCAGCAAGC----- 4963
QY 180 TyrLeuSerAlaSerProAspArgLeuHisAspAlaMetAlaAlaSerValLysArgHis 199
Db 4964 ---ACATCAGCTTCTGAATCTGCATCAACCAGTGCCTTCAGCTTCACCAAGCGCC 5020
QY 200 SerProSerLeuAlaArgGlnValLeuAspThrGlyValAlaValGlnThrTyrSerAla 219
Db 5021 TCGGCCCTCAGCAAGT-----ACAAGTGCCTTCAGCCTCAGCATCAACCAGT 5065
QY 220 ArgAsnAlaValArgThrValLeuAlaProAlaLeuAlaSerArgProAla--ValGln 238
Db 5066 GCATCAGCTTCAGCCTCAACAAGTGCCTTCAGCCTCAGCGTCAACCAAGTGCCTCGGTTCA 5125
QY 239 GlyAlaValAspLeuGlyValSerMetAlaGlyGlyLeuAlaAlaAsnAlaGlyPheGly 258
Db 5126 GCAAGTACCAGTGCCTCAGCTTCAGCAAGCAAGTGCCTTCAGCTTCAGCATCAACCAGT 5185
QY 259 Asn-ArgLeuLeuSerValGlnSerArgAspHisGlnArgGlyGlyAlaLeuValLeuGln 278
Db 5186 GCTTCGGCTTCGGCATCAACAAGTGCCTCAGCATCAGCAT----- 5225
QY 278 yLeuLysAspLysGluProLysAlaGlnLeuSerGluGluAsnAspTyrLeuGluAlaTyr 298
Db 5226 -CAACGAGTGCCTCAGCCTCAGCAAGTACTAGTCATCAGCATCAG-----CA 5272
QY 298 rLysAlaIleLeuSerAlaSerTyrSerGlyAlaAlaLeuAsnAlaGlyLysArgMetAl 318
Db 5273 TCAACCAAGTGA-TCAAGCCTCAACAAGTATCTCAGCGTCTGAATCGGCATCAACGAGTGC 5331
QY 318 aGlyLeuProLeuAspMetAlaThrAspAlaMetGlyAlaValArgSerLeuValSerAl 338
Db 5332 ATCAGCATCAGCATCAACGAGTGCATCGCTTCAGCGTCAACCAAGTGCATCAGTCTCAGC 5391
QY 338 aSer---SerLeuThrGlnAsnGlyLeuAlaLeuAlaGlyGlyPheAlaGlyValGlyLys 357
Db 5392 AAGCAACCAGTGCCTCGCTTCAGCATCAACGAGTGCCTCAGCGTTCAGCAAGTATCTCAGC 5451
QY 357 sleuGlnGluMetAlaThrLysAsnIleThrAspProAlaThrLysAlaAlaValSerGln 377
Db 5452 GTCTGAATCGCATCAACGAGTGCCTCAGCGTTCAGCAAGTACTAGTCATCGGCTTCAGC 5511

QY 377 nLeuThrAsnLeuAlaGlySerAlaAlaValPheAlaGlyTrrPThrAlaAlaLeuTh 397
Db 5512 AAGCAACAGTGCCTCGGCTTCAGCATCAACCAGTGCCTCA---GCCTCAGCAAGTACTC 5568
QY 397 rThrAspProAlaValLysLysAlaGluSerPheIleGlnAspThrValLysSerThrAl 417
Db 5569 AGCGTCTGAATCGGCATCAACGAGTGCCTCAGCCTCAGCAAGTACTAGTCATCAGCMTC 5628
QY 417 aSerSerThrThrGlyTyrValAlaAspGlnThrValLysLeuAlaLysThrValLysAs 437
Db 5629 AGCATCAACGAGTGCATCGGCTTCAGCAAGTACCAGCGCCTCAGCTTCAGCAAGCACAG 5688
QY 437 pMetGlyGlyGluAlaIleThrHisThrGlyAlaSerLeuArgAsnThr-ValAsnAsnL 457
Db 5689 TGGCTCAGCCTCAGCAAGTACCAGCGCCTCAGCTCA---GCAAGCACAGTGCCTCAGC 5745
QY 457 euArgGlnArgProAlaArgGluAlaAspIleGlu 469
Db 5746 TTCAGCAAGTACCAGTGCCTCAGCTCAGCATCAACAAG 5783

RESULT 7

US-08-961-527-364
; Sequence 364, Application US/08961527
; Patent No. 6420135
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 391
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,527
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 364:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2550 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-08-961-527-364

Alignment Scores:

Pred. No.: 0.000686 Length: 2550
Score: 135.50 Matches: 99
Percent Similarity: 37.02% Conservative: 85
Best Local Similarity: 19.92% Mismatches: 259
Query Match: 5.62% Indels: 55
DB: 4 Gaps: 11

US-09-825-414-7 (1-486) x US-08-961-527-364 (1-2550)

QY 2 HisileAsnArgArgValGlnGlnProProValThrAlaThrAspSerPheArgThrAla 21

D	b	815	CATCTGAATCAGCGT-----CAACCAGT-GCTTCGGCTTCAGCAAGTACCACTGCTTCA	867
O	y	22	SerAspAlaSerLeuAlaSerSerSerValArgSerValSerSerAspGlnIleArgGlu	41
D	b	868	GCTTCAGCATCAACCAAGCGCCCTCGCCCTCAGCAAGCACCTCAGCTTGTGAATCGGCCCTCA	927
O	y	42	IleAsnAlaIleAlaAspTyrLeuThrAspHisValPheAlaAlaHisIleLysLeuProPro	61
D	b	928	ACCAGCGCCCTCGGCCTCAGCAAGCACCTCAGCTTGTGAATCGGCCCTCAACCAGCGCCCTCA	987
O	y	62	AlaaspSerAlaAspGlyGlnAlaAlaValAlaAspValHisAsnAlaGlnIleThrAlaLeu	81
D	b	988	GCCTCAGCATCAACGAGTGCTTCGGCTTCAGCAAGCACAAAGCGCCTCGGGTTCAACATCA	1047
O	y	82	IleGluThrArgAlaSerArgLeuHisPheGluGlyGluThrProAlaThrIleAlaAsp	101
D	b	1048	ACGAGTAGCATCAGCTTCAGCGCTCAACCAAGTGTCTTCAGCCTCAGCATCAACAAGTGCCTCA	1107
O	y	102	ThrPheAlaLysAlaGluLysLeuAspArgLeuAlaThrThrTrpSerGlyAlaLeuArg	121
D	b	1108	GCCTCAGCAAGTAGTCTCAGCGTCTGAATCGGCATCAACGAGTGCCTTGAGTCAACATCA	1167
O	y	122	AlaThrProPheAlaMetAlaSerLeuLeuGlnTyrMetGlnProAlaIleAsnLysGly	141
D	b	1168	ACGAGTAGCATCAGCCCTCAGCAAGCACAGTAGTGTCTTCAGCCCTCAGCAAGTAGTCTCAGCGTCT	1227
O	y	142	AspTyrPleuProAlaProLeuLysProLeuThrProLeuIleSerGlyAlaLeuSerGly	161
D	b	1228	GAA-----TCGGCATCAACGAGTGCCTCCGCTTCAGCAAGTAGTAGC	1269
O	y	162	AlaMetAspGlnValGlyThrLysMetMetAspArgAlaThrGlyAspLeuHisTyrLeu	181
D	b	1270	GCCTCAGCATCAGCGTCAACAAAGTGCTTCGGCTTCAGCGTCA-----ACG	1314
O	y	182	SerAlaSerProAspArgLeuHisAspAlaMetAlaAlaSerValLysArgHisSerPro	201
D	b	1315	AGTGCCT-----GAGTCAGCAATCAACGAGTAGCTCAGCCTCAGCAAGC	1359
O	y	202	SerLeuAlaArgGlnValLeuAspThrGlyValAlaValGlnThrTyrSerAlaArgAsn	221
D	b	1360	ACATCAGCTTCTGAATCTGCATCAACCAAGTGCCTCAGCCTCAGCATCGACAAAGCCCTCA	1419
O	y	222	AlaValArgThrValLeuAlaProAlaLeuAlaSerArgProAla--ValGlnGlyAla	240
D	b	1420	GCTTCAGCAAGTAGTACCAAGTGCCTCAGCCTCAGCAAGTAGTGCTTCAGCCTCAGCGTGC	1479
O	y	241	ValAspLeuGlyValSerMetAlaGlyGlyLeuAlaAlaAsnAlaGlyPheGlyAsnArg	260
D	b	1480	ACAAGTGCCTCGCCCTCAACCAAGTGCATGTGAATCGGCATCAACCAAGTGCCTCAGCTCAG	1539
O	y	261	LeuLeuSerValGlnSerArgAspHisGlnArgGlyGlyAlaLeuValLeuGlyLeuLys	280
D	b	1540	CAAGTAGTAGTCATCAGCTTCAGCATCAACGAGTAGCATCGGCTT-----	1584
O	y	281	AspLysGluProLysAlaGln-LeuSerGluGluAsnAspTyrPleuGluAlaTyrLysAl	300
D	b	1585	CGGCGTCAACCAAGTGCATCAGAGTCAGCAAG-----TACCAGTGG	1623
O	y	300	allelysSerAlaSerTyrSergly-----AlaAlaLeuAsnAlaG1	314
D	b	1624	CGTCACTTCCGCATCAACAAGTGCCTCGGCTTCAGCAAGCACATCAGCATCTGAATCAGC	1683
O	y	314	yllysArgMetAlaGlyLeuProLeuAspMetAlaThrAspAlaMetGlyAlaValArgSe	334
D	b	1684	GTCAACCAAGTGTTCGGCTTCAGCAAGTAGTACAGTGTCTTCAGCATCAACCAAGCGC	1743
O	y	334	rLeuValSerAlaSer---SerLeuThrGlnAsnGlyLeuAlaLeuAlaGlyIlePheAl	353
D	b	1744	CTCGGCTCAGCAAGCAACTCAGCTTGTGAATCGGCCTCAACCAAGCGCCTCGGCTCAGC	1803
O	y	353	aglyValGlyLysLeuGlnGluMetAlaThrLysAsnIleThrAspProAlaThrLysAl	373

Db	1804	AAGCACTCAGCTTCTGTAATCGGCTCAACACAGCGCCCTCAGCCTCAGCATCAACGAGTGC	1863
QY	373	aalavalserglinleuthrasnleualaglyseralalavalpheaaglytrpthrth	393
Db	1864	TTCCGGCTTCAGCAACGACACAGCGCCTCGGGTTACGATCAACGAGT-----ACGTC	1914
QY	393	ralaalaleuthrthr-----aspbroal	401
Db	1915	AGCTTCAGCGTCAACAGCTGCTTCAGCCTCAGCATCAACAGTGCCTCAGCTCAGCAAG	1974
QY	401	avallyslsyalaguserpheiieglnaspthrvallysserthralaserserthrth	421
Db	1975	TATCTCAGCGTCTGAATCGCATCAACGAGTGCCTGTGAGTCAGCATCAACGAGTACGTC	2034
QY	421	rglytyrvalalalaspglnthrvallyslsleualalysthrvallysaspmetsglygl	441
Db	2035	AGCCTCAGCAACGACCTCAGCTTCTGAATCGGCCCTCAACCAAGTGCCTCAGCCTCAGCATC	2094
QY	441	ualaitlethrhisthrglyalaserleuargasnthrvalasnleuarglnargpr	461
Db	2095	GACACGCGCTCAGCTTCAGCAAGTACCAAGTGTTCAGCCTCAGCGTGCACAGTGCCTC	2154
QY	461	calaargluvalaspiieglnuglyglythrvalalaserproser	477
Db	2155	GGCCTCAACCAAGTCATCTGAATCGCATCAACCAAGTGCCTCAGCCTCA	2203
RESULT 8			
US-09-177-349-2	; Sequence 2, Application US/09177349		
	; Patent No. 6268201		
	; GENERAL INFORMATION:		
	; APPLICANT: Alland, David		
	; APPLICANT: Bloom, Barry R.		
	; APPLICANT: Jacobs Jr., William R.		
	; TITLE OF INVENTION: inh, inia AND inc GENES OF MYCOBACTERIA AND METHODS		
	; TITLE OF INVENTION: OF USE		
	; FILE REFERENCE: 96700/491		
	; CURRENT APPLICATION NUMBER: US/09/177,349		
	; CURRENT FILING DATE: 1998-10-23		
	; NUMBER OF SEQ ID NOS: 14		
	; SOFTWARE: Patentln Ver. 2.0		
	; SEQ ID NO 2		
	; LENGTH: 5036		
	; TYPE: DNA		
	; ORGANISM: Mycobacterium tuberculosis		
US-09-177-349-2			
Alignment Scores:			
Pred. No.:	0.00331	Length:	5036
Score:	133.50	Matches:	108
Percent Similarity:	33.41%	Conservative:	43
Best Local Similarity:	23.89%	Mismatches:	182
Query Match:	5.53%	Indels:	120
DB:	4	Gaps:	16
US-09-825-414-7 (1-486) x US-09-177-349-2 (1-5036)			
QY	64	Seralasp glylnalalavalaspyalhisasnala-----	76
Db	97	AGCGTCGATGGCACAGCGCGGTACGCCGCCGAACAGCTTCTTCGATTGAAGGAATG	156
QY	77	glnlethr alaleuileglu thr argalaser argleu hisphegluglylu thr pro	96
Db	157	AAGATGACCTCGCTTATCGATTACATCTGAGCCTGTTCGCGACGGAAGACGCCGCCGG	216
QY	97	AlathrileAlaaspthrPheAlaLysAlaGluLysLeuAspargLeuAlaThrThr	116
Db	217	TCGTTCTGTTGCC-----GCTCCGGACGGCCATGACCAGT	252
QY	117	SerglyAlaLeuArgAlaThrProPheAlaMetAlaSerLeuLeuGlnTyrMetGlnPro	136
Db	253	GCGGGCTGATCGATATCGCGCGCACCAATCTCATCGGTGGCGGCAATGTGTGCCG	312

QY 137 AlaIleasnlysgIyAspTrpLeuProAlaProLeuLysProLeuThrProLeuIleSer 156
Db 313 GGTCTGAATCTGGGTGCCGCGACCCC----- 339
QY 157 GlyAlaLeuSerGlyAlaMetAspGlnValGlyThrLys-MetMetAspArgAlaThrG1 176
Db 340 -----ATGAGCGGATTCGGGACAGCCCGCTCGCATGGCTTGGCAGAGACGTC 393
QY 176 yAspLeuHisTrpLeuSerAlaSerProAspArgLeuHisAspAlaMetAlaAlaSerVa 196
Db 394 GCCA-----ATGTCGGCTTCGGCGGTGACGGGGCGGGGGGTGGCAAGCGTCATC 444
QY 196 lLysArgHisSerProSer-----LeuAlaArgGlnValLeuAspThrGlyValAlaVa 214
Db 445 ACGACCGATGTCGGTCCGGCCCTGGCTAGCGGACTGGGTGCTGGTCTGGGTCAAGGT 504
QY 214 lGlnThrTrpSerAlaArgAsnAlaValArgThr----- 225
Db 505 GGCTGTGCTTCGCCCGCTCAAGCGGTGGTTTCGGCGGTCAAGTCCGGTTCGGCTGCCAG 564
QY 226 -----ValLeuAlaProAlaLeuAlaSerArgProAlaValGlnGlyAlaValas 242
Db 565 GTCGGTCTGGGTTTACTGCCGTGATG-----AGGCCGAG-GTCGGCGCTCAGGTTGG 617
QY 242 pLeuGlyValSerMetAlaGlyGlyLeuAlaAlaAsnAlaGlyPheGlyAsnArgLeuLe 262
Db 618 TGCTGGGTAGATATTGGGACGGGCTCGGGTCTCAGCCGGTATGGC----- 666
QY 262 uSerValGlnSerArgAspHisGlnArgGlyGlyAlaLeuValLeuGlyLeuLysAspLy 282
Db 667 -----TTTGGCGCGGGGGTGGCTGGCTGGGTCTGGGTGCT-- 699
QY 282 sGluProLysAlaGlnLeuSerGluGluAsnAspTrpLeuGluAlaTrpLysAlaIleLy 302
Db 700 -----CAGGCCGGCGGTGTGATCGG 719
QY 302 sSerAlaSerTrpSerGlyAlaAlaLeuAsnAlaGlyLysArgMetAlaGlyLeuProLe 322
Db 720 TGGAGCGCGCGCGGCTATCGGTGCCGCGGTGCGGCTAGCGGCAATGCG-- 777
QY 322 uAspMetAlaThrAspAlaMetGlyAlaValaArgSerLeuVal----- 336
Db 778 -CAGATCGAGATTGCCGGCCAGGGTGCCGTGGCGCTGTCCGGCGCTGTCCGGCGG 836
QY 337 -SerAlaSerSerLeuThrGlnAsnGlyLeuAlaLeuAlaGlyGlyPheAlaGlyValG1 356
Db 837 CCAGCGCGGCATCCCTAGCCAGATCGGTGTCTCAGCCGGTGGTGGCGCTCGCGCGCTCGG 896
QY 356 yLys----- 357
Db 897 CAATGTCAGCGCGCTGACCGGGGTCAAGCAAGCATGTTGGCTTCCAACGCAAGCGG 956
QY 358 -----LeuGlnGluMetAlaThrLysAsnI1 366
Db 957 CCAGGCGGGGTGATCGCCAGTGAAAGCGCGCTGCTTGAACGGCGCTGATGCTCATCT 1016
QY 366 eThrAspProAlaThrLysAlaAlaValSerGlnLeuThrAsnLeuAlaGlySerAlaAl 386
Db 1017 GTCGGCGCGCTTACCGCGGTGTCGTGGGTGTCAGGCGCGCGCTGGGCGCGCGG 1076
QY 386 aValPheAlaGlyTrpThrThrAlaAlaLeuThrThrAspProAlaValLysLysAlaG1 406
Db 1077 GTTGGGCTTCGGA---GCGGTCCGGCACCCGACTCTCAGCCGGCGCGCTGGCGCGG 1133
QY 406 uSerPheIleGlnAspThrValLysSerThrAlaSerSerThrThrGlyTrpValAlaAs 426
Db 1134 TGGCGTGTGGCCAGACCC-----GAGCGCGGTCTGAGTGGTGGCGGGGTTCGGCGG 1187
QY 426 pGlnThrValLysLeuAlaLysThrValLysAspMetGlyGlyGluAlaIleThrHisTh 446
Db 1188 GGCAACCC-----GCGGCGGGGTCCGGCGGCGACACGCGCATCTGGCGCCACGA 1238
QY 446 rGlyAlaSerLeuArg-----AsnThrValasn 455

Db 1239 GGGAGCCCGCACTGGCGAGTGTGCAGACACGGTCAAC 1272
RESULT 9
US-09-103-840A-2
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-2007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2
Alignment Scores:
Pred. No.: 247 Length: 4403765
Score: 132.50 Matches: 123
Percent Similarity: 34.23% Conservative: 55
Best Local Similarity: 23.65% Mismatches: 195
Query Match: 5.49% Indels: 147
Gaps: 21
US-09-825-414-7 (1-486) x US-09-103-840A-2 (1-4403765)
QY 20 ThrAlaSerAspAlaSerLeuAlaSerSerSerValArgSerValSer----- 35
Db 1632092 ACCGCCCTGCCCGCGCGCGCGGGGTACCAGCGGTACCAGCGCGCGGTACCATC 1632151
QY 36 SerAspGlnGlnArgGluLeuAsnAlaAlaIleAlaAspTrpLeuThrAspHisValPheAla 55
Db 1632152 GCCGAAGCGGAAGTTCGCTTGGCGCGGTGCGCGCGGTACACCGCGGAGCCGCGCGCC 1632211
QY 56 AlaHisLysLeuProProAlaAspSerAlaAspGlyGlnAlaAlaValaAspValHisAsn 75
Db 1632212 CCTTTGCCGCGCGGACCGCGGACACACCTGTGATCCGTTCTGGCCAAAGAGTTCCCGGC 1632271
QY 76 -----AlaGlnIleThrAlaLeuIleGluThrArgAlaSerArg 88
Db 1632272 CAACACGCGCGCGCGCTTGGCGCGCGGTACCAGCTTG-----CGCGCC----- 1632316
QY 89 LeuHisPheGluGlyLutThrProAlaThrIleAlaAspThrPheAlaLysAlaGluLys 108
Db 1632317 -----GGGCGCGCGCGGTGGCGCGCGCGCGCGCTTGGCGCGCGCGGTGGC 1632364
QY 109 LeuAspArgLeu-----AlaThrThrThrSerGlyAlaLeuArgAlaThrPro 124
Db 1632365 GTTGGCGCGCGGTGGCGCGGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1632424
QY 125 PheAlaMetAlaSerLeuLeuGlnTrpMetGlnProAlaIleasnlysgIyAspTrpLeu 144
Db 1632425 ACCGCTCCCGCGCGCGCGCGCGAAGTCTCCGCGCGCGC----- 1632463
QY 145 ProAlaProLeuLysProLeuThrProLeuIleSerGlyAlaLeuSerGlyAlaMetAsp 164
Db 1632464 GCCGCGCGCGCGCGCGCGCGCAAGCCCGCGGTTCGCGCGGTTCGCGCGGATTCGCGCG 1632523
QY 165 GlnValGlyThrLysMetMetAspArgAlaThrGlyAspLeuHisTrpLeuSerAlaSer 184

Db 1632524 GGTGGTGGG-----CGAGGAACCGGC----- 1632544
QY 185 ProAspArgLeuHisAspAlaMetAlaAlaSerValLysArgHisSerProSerLeuAla 204
Db 1632544 ----- 1632544
QY 205 ArgGlnValLeuAspThrGlyValAlaValGlnThrTyrSerAlaArgAsnAlaValArg 224
Db 1632545 -----GACACCGCCCATGCGCTGCCCGCTTGGCCGC----- 1632577
QY 225 ThrValLeuAlaProAlaLeuAlaSerArgProAlaValGlnGly----- 239
Db 1632578 -----GGCCCCCGCCATACCAACAAGCCCGCCGCTTGGCCGCTGGCCCC 1632628
QY 240 -----AlaValAspLeuGlyValSer-----MetAlaGlyGlyLeuAlaAlaAsn 254
Db 1632629 GCCGGCCCCCGCCGGCAGCGGTGGCGCTTGGCGCGTGGCCGGTGGCCGCC 1632688
QY 255 AlaGlyPheGlyAsnArgLeuLeuSerValGlnSerArgAspHisGlnArgGlyGlyAla 274
Db 1632689 GCTGTCGGGGTGGCGCCGCGACACCGTCTGCACCCCGCGGTGATCC--GGCGCCGCC 1632745
QY 275 LeuValLeuGlyLeuLysAspIysGluProLysAlaGlnLeuSerGluGluAsnAspTrp 294
Db 1632746 GATCCCAACCAACACACACCGATGCGCGGCTACCGCGCTTGGCGCGCTTGGCAACCAACTCC 1632805
QY 295 LeuGlnAlaTyr-----LysAlaIleLysSerAlaSerTyrSer 307
Db 1632806 ATCGCCCGCGCTTATCGAACGTGCCCTTGGCACCGTTCGCCGCATCACCGCCATGCGCGCC 1632865
QY 308 GlyAlaAlaLeuAsnAlaGly----- 314
Db 1632866 GCGCGCGCGCTTCCGCGCGCGCGCGCGCACCCATGCTGCCGCTTGGTGGGTGGCTGC 1632925
QY 315 LysArgMetAlaGlyLeuProLeuAspMetAlaThrAspAlaMetGlyAlaValArgSer 334
Db 1632926 AAGCGCCTTACCGCCTTGCCACCGGCTCCACCGCACCGCGGCTCCACCGTGGCGGCC 1632985
QY 335 LeuValSerAlaSerSerLeuThrGlnAsnGlyLeuAlaLeuAlaGlyGlyPheAlaGly 354
Db 1632986 CTTGCCCGCGCTGCGTGCATCCGCGCGCTGCCCGCAGCGCGCTTAAGCGGTGGCGCGGT 1633045
QY 355 ValGlyLysLeuGlnGluMetAlaThrLysAsnIleThrAspProAlaThrLysAlaAla 374
Db 1633046 -----GGCGCGCTTGGCGGCC 1633060
QY 375 ValSerGlnLeuThrAsnLeuAlaGlySerAlaAlaValPheAla--GlyTrpThrThr 393
Db 1633061 GTTGGCGCGCTTACCGCGCGCGCGCGCACACCGCGCTCGCCTGCTTGGGCTCCGCCGTC 1633120
QY 394 AlaAlaLeuThrThrAspProAlaValLysLysAlaGluSerPhe----- 408
Db 1633121 GCCGCCCTTACCGCCAGCGCGCGCACCTCCGCCGCCACCGCGCTTAGGGTCGCCGCCAGA 1633180
QY 409 IleGlnAspThrValLysSerThrAlaSerSerThrThrGly--TyrValAlaAspGln 427
Db 1633181 AGCGGGGGCACCGGGGGCGCGCTTGGCGCGCGCGCGCATTTGCCGACACAG 1633240
QY 428 ThrValLysLeuAlaLysThrValLysAspMetGlyGlyGluAlaIleThrHisThrGly 447
Db 1633241 CCCGCCGCCCGCGC-----GGCGCGCGCTTACC--GCCGCGC 1633276
QY 448 AlaSerLeuArgAsnThrValAsnAsnLeuArgGlnArgProAla----- 462
Db 1633277 TTTGGCGCGCGATGAGAAGTGGCGCGCTTGGCGCGCGCGCGCTTGGCGCGCTGGT 1633336
QY 463 -----ArgGlnAlaAspIleGluGluGlyGlyThrAlaAlaSerProSer 477
Db 1633337 GGGGCTGGCCCCCGCGCGCGCGCGCATCTGTGAGCGCGCTCCGCCGTTGCCCTCC 1633396
RESULT 10
US-08-804-227C-7
; Sequence 7, Application US/08804227C

; Patent No. 5876991
; GENERAL INFORMATION:
; APPLICANT: DeHoff, Bradley S.
; APPLICANT: Kuhstoss, Stuart A.
; APPLICANT: Rostock, Paul R., Jr.
; APPLICANT: Sutton, Kimberly L.
; TITLE OF INVENTION: POLYKETIDE SYNTHASE GENES
; NUMBER OF SEQUENCES: 15
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: THOMAS G. PLANT 1501
; STREET: LILLY CORPORATE CENTER
; CITY: INDIANAPOLIS
; STATE: IN
; COUNTRY: USA
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: ASCII(DOS) Text only
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/804,227C
; FILING DATE: February 21, 1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Plant, Thomas, G.
; REGISTRATION NUMBER: 35,784
; REFERENCE/DOCKET NUMBER: X-8231
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317-276-2459
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 44377 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 350..14002
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 14046..20036
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 20110..31284
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 31329..36071
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 36155..41830
; US-08-804-227C-7
Alignment Scores:
Pred. No.: 0.228 Length: 44377
Score: 130.50 Matches: 135
Percent Similarity: 33.77% Conservative: 73
Best Local Similarity: 21.92% Mismatches: 214
Query Match: 5.41% Indels: 195
DB: 2 Gaps: 28
US-09-825-414-7 (1-486) x US-08-804-227C-7 (1-44377)
QY 12 ValThrAlaThrAspSerPheArgThrAlaSerAspAlaSerLeuAlaSerSerVal 31
Db 18069 CTGACCAACACCGCAGCGCGCGTACGGCGTCGAGCGGGGG--CGGCTCAGTCTCTC 18125
QY 32 ArgSerValSerSerAspGlnGlnArgGluIleAsnAlaIleAlaAspTyrLeuThrAsp 51
Db 18126 GACGAGGTGCGTGCAGCGGTGCGCGCGGACCACTCGAAGCCGTTACAGCCCGCTCGCCGAC 18185
QY 52 HisVal-----PheAlaAlaHis 57

Db 18186 ACCGAAACGGCCCCGACGTCGACGTCCGTACGGCCCCGCCCCCGCACAGCCGCCGT 18245
QY 58 LysLeuProProAlaAspSerAlaAspGlyGlnAlaValAlaValHisAsnAlaGln 77
Db 18246 ACAAGCCCCCGGTGGACACCCGACGGGAGGCCCGCACCGCTGAC-----GGCCCCGG 18299
QY 78 IleThrAlaLeuIleGluThrArgAlaSerArgLeuHisPheGluGly-----GluThr 95
Db 18300 CTCGTCTCTGACCCCGGGCGCGCCGAGCC-----GAGGAGCGCGCGCGAT 18350
QY 96 ProAlaThrIleAla-----AspThrPheAlaLysAlaGluLysLeuAspArg 111
Db 18351 CCGCGCGGTCCGCTGTCTGGGGCTCTCGTCCGGTCCGCCAGCCGGAACAGCCCGCGCGC 18410
QY 112 LeuAlaThrThr-----ThrSerGlyAlaLeuArgAlaThrProPheAlaMet 127
Db 18411 TTCACCTGTGTGACGTGACGGGACCCAGCGCTCGCTGGGGCCCTGCCCGGTCTGCTG 18470
QY 128 AlaSerLeuLeuGlnTyrMetGlnProAlaIleAsnLysGlyAspTrpLeuProAlaPro 147
Db 18471 GCCACG-----GATGCCGCGCAGTCCGGCGGTGCGCGACGACGTTCACCGTCCCGCGC 18524
QY 148 LeuLysProLeuThrProLeuIle----- 155
Db 18525 CTCGTCTCCGGTGGCGGACCCCGCTCCCGCACGGCGGCGGCGCGCGCGGACGGGT 18584
QY 156 SerGlyAlaLeuSerGlyAlaMetAspGlnValGlyThrLysMetMetAspArgAlaThr 175
Db 18585 GCCGGCGAGCCGTCTCCGCGACCCCTGGACCCCGAAGGACACCTGCTGATCACCGCGCGCAC 18644
QY 176 GlyAspLeuHisTyrLeuSerAlaSerProAspArgLeuHisAspAlaMetAlaLaser 195
Db 18645 GGA-----GCACTGGCCGCGGAA 18662
QY 196 ValLysArgHis----- 199
Db 18663 ACCGCGCGGACCTGTGTCAGCCGACACAAGTGGCCCATCTCTGCTGTGGGACAGGCGC 18722
QY 200 -----SerProSerLeuAlaArgGlnValLeuAsp-----ThrGly 211
Db 18723 GGTCCCGACGACACCCGCGCTCATGACTGCTGCCCGAGTGTGACCCGAGTCCGGTCCGAG 18782
QY 212 ValAlaValGlnThr-----TyrSerAlaArgAsnAlaValArgThrValLeuAlaPro 229
Db 18783 GTCCGCTGACGGGCTGTGACTGACGGAGCCGCGACGCCCTGCGCGCGCTGCTGACGCA 18842
QY 230 AlaLeuAlaSerArgProAla-----ValGlnGlyAlaValAspLeuGly 244
Db 18843 CTCGCCGACGAACACCCGCTGACTGCTGTCACACCCGCGGGGTGCTCGACGAGGCG 18902
QY 245 ValSerMetAlaGlyGlyLeuAlaAlaAsnAlaGlyPheGlyAsnArgLeuLeuSerVal 264
Db 18903 GTG-----CTCTCCGCC 18914
QY 265 GlnSerArgAspHisGlnArgGlyAlaLeuValLeuGlyLeuLysAspLysGluPro 284
Db 18915 CAGACGCGCGGAGCG-GATCGACACGGTGTCTCCGCGCCAAAGCGGACGCGCGCTCCACT 18973
QY 285 ---LysAlaGlnLeuSerGluLysAsnAspTrp-LeuGluAlaTyrLysAlaIleLysSe 303
Db 18974 GACGAGCTGACCCGGAGATCGAGCGGGTGCCTGCTGTACTCTCGTCTCGGC 19033
QY 303 r-----AlaSerTyrSerGlyAla-----AlaLe 311
Db 19034 CACCTGTGGCAGCGCGGGGACAGCCGGGTACGCGGCGGCCAACGCTTCATGACGCGCT 19093
QY 311 uAsnAlaGlyLysArgMetAlaGlyLeuPro----- 321
Db 19094 GCGCGCGCGCGGTGCGCGCGCGGACACCCGCGCTGCTGCGCTGGGCTGGGTGCTG 19153
QY 322 ---LeuAspMetAlaThrAspAlaMetGlyAlaValArgSerLeuValSerAlaSerSe 340
::: ||||| |||||

Db 19154 CCGGCTGGGTCTCGCCACCGGAGCTGACGGAGCG-----GACGCGCGCG 19198
QY 340 rLeuThrGlnAsnGlyLeuAlaLeuAlaGlyGlyPheAlaGlyValGlyLysLeuGlnI 360
Db 19199 GGTCAAGCGCTCGGCTCTCGCCCGCTCGACGCGCGCGCGGACCTGACCTGTCGACCG 19258
QY 360 uMetAlaThrLysAsnIleThrAspProAlaThrLysAlaAlaValSerGlnLeuThrAs 380
Db 19259 GCGCGTACCCCG-----CCCGAGCCGCGCTGCTGCCCGTCCGCGCTGACCTGCGCGC 19312
QY 380 nLeuAlaGlySerAlaAlaVal-----PheAlaGlyTrpThrAl 394
Db 19313 CCGCGCGGTGCGCACCGCTCTCCGGAGGTCTGCTGCTGACCTGCGCGCGCGCGGA 19372
QY 394 aAlaLeuThrThr----- 404
Db 19373 CGCCCGACGACGCGCGCGCGCGCGCGGACCGGAGAGAGACGAGTGCCTGCGCGC 19432
QY 404 sAla----- 405
Db 19433 TGCCCCCGCCCCGCGGACGCGCGCGGAGCGCTGGCGCGCGCGCTCGCGGACGTTCCGC 19492
QY 406 -----GluSerPheIleGlnAspThrValLysSerThrAlaSerSerThrThrGl 422
Db 19493 ACCCGAGCGTACGCGCTCTCTCGACCTGCTGCGGACCGAGTCCGCGCGGTGCTCGG 19552
QY 422 yTyrValAlaAspGlnThrValLysLeuAlaLysThrValLysAspMetGlyGlyGluAl 442
Db 19553 ACACGCGGACCCCGCGCGGATCGCGCGCGCGCGCGCACCTTCAGAGACCGCGGATTCGACTC 19612
QY 442 aIleThrHisThrGlyAlaSerLeuArgAsnThrValAsn-----AsnLeuArg-- 458
Db 19613 CCTC-----ACCGTGTGACCTTCGCAACCGGCTGAACACAGCACCGGACTGCGGCT 19666
QY 459 -----GlnArgProAlaArgGluAlaAsp----- 466
Db 19667 GCCCGGACCTCTGTTTCAGACACCCACACCGCTCGCGCTCGCGAAGCTCTGCTGA 19726
QY 467 ---IleGluGluGlyGlyThrAlaAlaLaserProSerGluIlePro 480
Db 19727 CCGGCTGGAGGCGCGCGGTCCAGCGGAACCGCGCGCTGAGTTCGCG 19772

RESULT 11
US-08-804-198-1
; Sequence 1, Application US/08804198
; Patent No. 5945320
; GENERAL INFORMATION:
; APPLICANT: Burgett, Stanley G.
; APPLICANT: Kuhstoss, Stuart A.
; APPLICANT: Rao, Nagaraja R.
; APPLICANT: Richardson, Mark A.
; APPLICANT: Rostock, Paul R., Jr.
; TITLE OF INVENTION: PLATENOLIDE SYNTHASE GENE
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: PAUL R. CANTRELL 1138
; STREET: LILLY CORPORATE CENTER
; CITY: INDIANAPOLIS
; STATE: IN
; COUNTRY: USA
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: Macintosh
; OPERATING SYSTEM: Macintosh 7.0
; SOFTWARE: Microsoft Word 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/804,198
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: CANTRELL, PAUL R.
; REGISTRATION NUMBER: 36,470

```

; REFERENCE/DOCKET NUMBER: P9113
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317-276-3885
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 44377 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 350..14002
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 14046..20036
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 20110..31284
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 31329..36071
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 36155..41830
; US-08-804-198-1

Alignment Scores:
Pred. No.: 0.228
Score: 130.50
Percent Similarity: 33.77%
Best Local Similarity: 21.92%
Query Match: 5.41%

Length: 44377
Matches: 135
Conservative: 73
Mismatches: 214
Indels: 195
Gaps: 28

US-09-825-414-7 (1-486) x US-08-804-198-1 (1-44377)
QY 12 ValThrAlaThrAspSerPheArgThrAlaSerAspAlaSerLeuAlaSerSerVal 31
Db 18069 CTGACCACACCGGACCGCGCGCTACGGCGTGGAGGGGG--CGGCTCAGTCTCTC 18125
QY 32 ArgSerValSerSerAspGlnArgGluIleAsnAlaIleAlaAspTyrLeuThrAsp 51
Db 18126 GACGAGTGGCTGACGAGTGGCGGACCGACCTCGAAGCCGTACAGGCCCGCTCGCGAC 18185
QY 52 HisVal-----PheAlaIleHis 57
Db 18186 ACCGAACGCGCCCGAGCTGACGTCGTACGGCGCGCGCCCGCCGACAGCGCGCT 18245
QY 58 LysLeuProProAlaAspSerAlaAspGlyGlnAlaIleValAspValHisAsnIleGln 77
Db 18246 ACAAGCCCCCGGTGACACCCGACGAGGAGCGCCGACCGCTGAC-----GGCCCCGG 18299
QY 78 IleThrAlaLeuIleGluThrArgAlaSerArgLeuHisPheGluGly-----GluThr 95
Db 18300 CTCGTCGTCGTGACCGCGGGCGCGCGGACCC-----GAGGAGGGCGGCGCAT 18350
QY 96 ProAlaThrIleAla-----AspThrPheAlaLysAlaGluLysLeuAspArg 111
Db 18351 CCGCGGGGTGCGGTGTGTGGGGCTGTCGCGGGTCCGCAAGCGCAAGCCGCGCGC 18410
QY 112 LeuAlaThrThr-----ThrSerGlyAlaLeuArgAlaThrProPheAlaMet 127
Db 18411 TTCACCCCTGGTGAAGTGAAGGACCCAGAGCGTCTGCGGGCCCTGCGCGTCTGCTG 18470
QY 128 AlaSerLeuLeuGlnTyrMetGlnProAlaIleAsnLysGlyAspTyrLeuProAlaPro 147
Db 18471 GCCACG-----GATGCGGGCCAGTGGCGCTGCGGACGAGCGTGTACCGCTCCGCGC 18524
QY 148 LeuLysProLeuThrProLeuIle----- 155
Db 18525 CTCGTCCCGGTGGCGGACCCCGTCCCGCACGGCGGCGGACGCGCGGCGGAGCGGCT 18584
```

```

QY 156 SerGlyAlaLeuSerGlyAlaMetAspGlnValGlyThrLysMetMetAspArgAlaThr 175
Db 18585 GCGGCGAGCCGCTCCGCGACCTGGACCCCGAAGGACCGCTGATCACCAGCGGACCC 18644
QY 176 GlyAspLeuHisTyrLeuSerAlaSerProAspArgLeuHisAspAlaMetAlaSer 195
Db 18645 GGA-----GCACGTGCGCGGGA 18662
QY 196 ValLysArgHis----- 199
Db 18663 ACCGCGCGGACCTGTCGACCGGACACAGAGTGCGCCATCTCTGCTGCGGACGCGC 18722
QY 200 -----SerProSerLeuAlaArgGlnValLeuAsp-----ThrGly 211
Db 18723 GGTCCCGACGACCGCGCGCTCGATCGACTGTCGCCGAGTTGACCGAGTCCGGTCCCGAG 18782
QY 212 ValAlaValGlnThr-----TyrSerAlaArgAsnAlaValArgThrValLeuAlaPro 229
Db 18783 GTCCCGGTACGGCGCTGTGACGTACGAGACCGCGCGCCCTGCGCGCTGCTCGACCGCA 18842
QY 230 AlaLeuAlaSerArgProAla-----ValGlnGlyAlaValAspLeuGly 244
Db 18843 CTCCCGGACGACACCGCTGACCTGCTGCTGTCACACCGCGGGGTGCTGACGACGCGC 18902
QY 245 ValSerMetAlaGlyGlyLeuAlaIleAsnAlaGlyPheGlyAsnArgLeuLeuSerVal 264
Db 18903 GTG-----CTCTCCGCGC 18914
QY 265 GlnSerArgAspHisGlnArgGlyGlyAlaLeuValLeuGlyLeuLysAspLysGluPro 284
Db 18915 CAGACGCGCGAGCG-GATCGACACAGGTGCTCGGCGCAAGCGGACGCGCGCTCCACCT 18973
QY 285 ---LysAlaGlnLeuSerGluGluAsnAspTyr-LeuGluAlaTyrLysAlaIleLysSe 303
Db 18974 GAGCAGAGCTGACCCCGGAGATCGGACGGGTGCCCTGTGCTGTACTCTCGCTCTCGGC 19033
QY 303 r-----AlaSerTyrSerGlyAla-----AlaLe 311
Db 19034 CACCCTGGGACGCGCGGCGGACGCGCGGTATCGCGCGGCAACGCTTCATGACGCGCT 19093
QY 311 uAsnAlaGlyLysArgMetAlaGlyLeuPro----- 321
Db 19094 GGCGCGCGCGGCTGCGCGCGCGCGGACACCGCGCTGCTCGGCTGGGCTGTGCTC 19153
QY 322 ---LeuAspMetAlaThrAspAlaMetGlyAlaValArgSerLeuValSerAlaSerSe 340
Db 19154 CGGGGTGGTCTCGCCACCGGACTGAGCGGAGCG-----GACGCGGCGCG 19198
QY 340 rLeuThrGlnAsnGlyLeuAlaLeuAlaGlyGlyPheAlaGlyValGlyLysLeuGlnI 360
Db 19199 GGTCAAGCGCTCGGTCTCGCGCGCGCTCGACGCGCGCGCGACCTGACCTGTGACCG 19258
QY 360 uMetAlaThrLysAsnIleThrAspProAlaThrLysAlaAlaValSerGlnLeuThrAs 380
Db 19259 GCGGCTGACCGG-----CCCGAGCGCGCTGCTGCCGTGCGCTGACCTGCGCGC 19312
QY 380 nLeuAlaGlySerAlaIleVal-----PheAlaGlyTyrThrThrAl 394
Db 19313 CGCGGCGCGGTGCCACCGCTCTCCGGAGGTCTCTGACCTGCGCGCGCTACCGGCGGA 19372
QY 394 aAlaLeuThrThr-----AspProAlaValLysLys 404
Db 19373 CGCCCGGACAGCGCGCGGCGCGCGGCGGACCGGGAAGAGACGAGTGCCTGCGCGCC 19432
QY 404 sala----- 405
Db 19433 TGCCCGCGCGCGCGGACGCGCGGAGCGTGGCGCGGAGCTTCGCC 19492
QY 406 -----GluSerPheIleGlnAspThrValLysSerThrAlaSerSerThrThrG 422
Db 19493 ACCCGAGCGTACGGCTCTCTGCTCGACCTGCTGCGGAGACGAGTGCAGCGGCTCTCG 19552
QY 422 yTyrValAlaAspGlnThrValLysLeuAlaLysThrValLysAspMetGlyGlyGluAl 442
```


Db 19553 ACACGGCAGCCCCCGCGATCGCGCGCCGCCACCTTCAAGGACGCCGGAFTCGACTC 19612
QY 442 aIeThrHsrHrGlyAlaSerLeuArgAsnThrValAsn-----AsnLeuArg-- 458
Db 19613 CCTC-----ACCGCTGTGACACTCCGCCAACCGGCTGAACACACGACCGGACTGCGGCT 19666
QY 459 -----GlnArgProAlaArgGluAlaAsp----- 466
Db 19667 GCCCGGACCCCTCGTCTTCGACCACCCACACCGCTCGCCCTCGCGGAATCTCTGCTCGA 19726
QY 467 ---IleGluGlyGlyThrAlaAlaSerProSerGluIlePro 480
Db 19727 CGGGCTGAGGCGCGCGTCCAGCGGAACCGCGCGCTGAGTCCCG 19772
RESULT 12
US-08-961-527-363
; Sequence 363, Application US/08961527
; Patent No. 6420135
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 391
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 Inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,527
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 363:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4483 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; US-08-961-527-363
Alignment Scores:
Pred. No.: 0.00788 Length: 4483
Score: 129.00 Matches: 96
Percent Similarity: 35.98% Conservative: 81
Best Local Similarity: 19.51% Mismatches: 269
Query Match: 5.35% Indels: 48
DB: 4 Gaps: 8
US-09-825-414-7 (1-486) x US-08-961-527-363 (1-4483)
QY 5 ArgArgValGlnGlnProProValThrAlaThrAspSerPheArgThrAlaSerAspAla 24
Db 1382 CGACAAGTGGCGGCTCA-ACCAGTGCATCTGAATCGGCATCAACACAGTGCCTCAGCC 1440
QY 25 SerLeuAlaSerSerSerValArgSerValSerSerAspGlnGlnArgGluIleAsnAla 44

Db 1441 TCAGCAAGTACTAGCGCCTCAGCCTCAGCATCAACG-----AGTGGC 1482
QY 45 IleAlaAspTyrIleuThrAspHisValPheAlaAlaHisLysLeuProProAlaAspSer 64
Db 1483 TTCGGCTTCAGCAAGTACTAGTGCATCAGCTTCGACCAAGTACTAGCGCCTCAGCCTCAGCG 1542
QY 65 AlaAspGlyGlnAlaAlaValAspValHisAsnAlaGlnIleThrAlaLeuIleGluThr 84
Db 1543 TCGACAAGCGCCTCAGCTTCAGCAAGTACCAGTGCCTCAGCCTCAGCGTCGACAAAGTGGC 1602
QY 85 ArgAlaSerArgLeuHisPheGluGlyGluThrProAlaThrIleAlaAspThrPheAla 104
Db 1603 TCGGCTTCAGCAAGTACTCAGCGTCTGAATCAGCATCAACAAGTGCCTGCGCTTCAGCA 1662
QY 105 LysAlaGluLysLeuAspArgLeuAlaThrThrThrSerGlyAlaLeuArgAlaThrPro 124
Db 1663 TCAACGAGTGCATCAGCTTCAGCATCAACAAGTGCCTTCAGCTTCAGCAAGTACCAGTGGC 1722
QY 125 PheAlaMetAlaSerLeuLeuGlnTyrMetGlnProAlaIleAsnLysGlyAspTyrLeu 144
Db 1723 TCGGCTTCAGCATCAACGAGTGCCTTCAGTTCAGCGTCAACACAGTGCCTCTGA----- 1776
QY 145 ProAlaProLeuLysProLeuThrProLeuIleSerGlyAlaLeuSerGlyAlaMetAsp 164
Db 1777 -----TCCGCATCAACAAGTGCCTTCAGCAAGCAGTGCCTTCGCGCT 1824
QY 165 GlnValGlyThrLysMetMetAspArgAlaThrGlyAspLeuHisTyrLeuSerAlaSer 184
Db 1825 TCAGCGTCAACGAGTGCCTGAGTCAAGTCA-----ACGAGTGCCTCA 1869
QY 185 ProAspArgLeuHisAspAlaMet----- 192
Db 1870 CCTCAGCAAGCAGCATCAGCTTCTGAATCTGCATCAACACAGTGCCTCAGCTTCGCGATCAAC 1929
QY 193 -AlaAlaSerValLysArgHisSerProSerLeuAlaArgGlnValLeuAspThrGlyVa 212
Db 1930 AAGCGCCTCGGCTCAGCAAGTACAGTGCCTTCAGCCTCAGCATCAACACAGTGCATCAGC 1989
QY 212 LAlaValGlnThrTyrSerAlaArgAsnAlaValArgThrValLeuAlaProAlaLeuAl 232
Db 1990 TTCAGCCTCAACAGATGCTTCAGCCTCAGCGTCAACCAAGTGCCTCGGCTTCAGCAAGTAC 2049
QY 232 aSerArgProAla-----ValGlnGlyAlaValAspLeuGlyValSerMe 247
Db 2050 CAGTGCCTCAGCTTCAGCAAGCAGCAGTGCCTCAGCTTCAGCATCAACCAAGTGCCTTCGGC 2109
QY 247 tAlaGlyGlyLeuAlaAlaAsnAlaGlyPheGlyAsnArgLeuLeuSerValGlnSerAr 267
Db 2110 TTCGGCATCAACAGTGCCTCAGCATCAGCATCAACGAGTGCCTCAGCCTCAGCAAGTACT 2169
QY 267 gAspHisGlnArgGlyGlyAlaLeuValLeuGlyLeuLysAspLysGluProLysAlaG1 287
Db 2170 AGTGCATCAGCAT-----CAGCATCAACCAAGTGCATCA 2202
QY 287 nLeuSerGluGluAsnAspTyrPleuGluAlaTyrLysAlaIleLysSerAlaSerTyrSe 307
Db 2203 GCCTCAGCAAGTATCTCAGCGTCTGAATCGGCATCAACGAGTGCA-TCAGCATCAGCATC 2261
QY 307 rGlyAlaAlaLeuAsnAlaGlyLysArgMetAlaGlyLeuProLeuAspMetAlaThrAs 327
Db 2262 AACGAGTGCATCGGCTTCAGCGTCAACCAAGTGCATCAGTCTCAGCAACCAAGTGCCTC 2321
QY 327 pAlaMetGlyAlaValArgSerLeuValSerAlaSer---SerLeuThrGlnAsnGlyLe 346
Db 2322 GGCTTCAGCATCAACAGTGCCTCAGCCTCAGCAAGTATCTCAGCGTCTGAATCGGCATC 2381
QY 346 uAlaLeuAlaGlyGlyPheAlaGlyValGlyLysLeuGlnGlnMetAlaThrLysAsnI1 366
Db 2382 AACGAGTGCCTCAGCGCTCAGCAAGTACTAGTGCATCAGCATCAGCATCAACGAGTGCATC 2441
QY 366 eThrAspProAlaThrLysAlaAlaValSerGlnLeuThrAsnLeuAlaGlySerAlaAl 386
Db 2442 GGCTTCAGCAAGTACCAAGCGCCTCAGCTTCAGCAAGCAGCAGTGCCTCAGCATCAGCAAG 2501

Db 21084 TGATGTACACCGCGCGG-AGAGGTCGGGCTGCGGCCAGTAGCGCTCGGTCTGGAAGCG 21026
QY 346 uAlaLeuAla-----GlyGlyPheAlaGlyValGlyLysLeuGlnGluMetAlaThr-- 363
Db 21025 TAGGTGGCAGATCGGGGTGGGTGCTGTG-GTGGGAGAGAGAGAGCCAGTCGACGGT 20967
QY 364 -----LysAsnIleThrAspProAlaThr 371
Db 20966 GAGGCGGTGGCCAGCGCTCGGCGAGGAGGTGTGAGCGCGGTCTGTCCGCGTGTGTC 20907
QY 371 rLysAlaAlaValSerGlnLeuThrAsnLeuAlaGlySerAlaAlaValPheAlaGlyTr 391
Db 20906 ACGCGGAGGAGGTGCCGAGCGGTGACGGTCTCGGGCAGGCCCATGTGAGAGCGGGGTG 20847
QY 391 pThrThrAlaAlaLeuThrThrAspProAlaValLys 403
Db 20846 G-----GCGCTGACCTCGACGAAGTGGGTGAAG 20819

RESULT 14
US-09-147-236-1
; Sequence 1, Application US/09147236A
; Patent No. 6316251
; GENERAL INFORMATION:
; APPLICANT: TONOUCHI, Naoto
; APPLICANT: TSUCHIDA, Takayasu
; APPLICANT: YOSHINAGA, Fumihiko
; APPLICANT: TAHARA, Naoki
; APPLICANT: HAYASHI, Takahisa
; TITLE OF INVENTION: NOVEL GENE, GROUP OF GENES, AND NOVEL BETA-GLUCOSIDASE
; FILE REFERENCE: 6537-011-OPCT
; CURRENT APPLICATION NUMBER: US/09/147,236A
; CURRENT FILING DATE: 1999-04-08
; EARLIER APPLICATION NUMBER: PCT/JP97/03633
; EARLIER FILING DATE: 1997-10-09
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 16836
; TYPE: DNA
; ORGANISM: Acetobacter xylinum
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (869)..(1891)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (3101)..(5368)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (5373)..(7778)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (7784)..(11761)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (11764)..(12231)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (12448)..(14652)
; FEATURE:
; OTHER INFORMATION: n at positions 15741 and 15767 may be a, g, t, or
; OTHER INFORMATION: c
US-09-147-236-1

Alignment Scores:
Pred. No.: 0.0852
Score: 128.00
Percent Similarity: 39.95%
Best Local Similarity: 25.20%
Query Match: 5.31%
DB: 4
Length: 16836
Matches: 94
Conservative: 55
Mismatches: 163
Indels: 62
Gaps: 17

US-09-825-414-7 (1-486) x US-09-147-236-1 (1-16836)

QY 10 ProProValThrAlaThrAspSerPheArgThrAlaSerAspAlaSerLeuAlaSerSer 29
Db 5535 CCGCCCGCCACAACC-----CCGGCGGGGATGCCCTCAGCAGCCAGCCGCG 5579
QY 30 SerValArgSerValSerSerAspGlnGlnArgGluIleAsnAlaIleAla-----Asp 47
Db 5580 GGTGATCGGGTGTGTGACAATGCCGAG-----AACGCCATCGCGGGGTGTGAC 5627
QY 48 TyrLeuThrAspHisValPheAlaAlaHisLysLeuProProAlaAspSerAlaAspGly 67
Db 5628 GTGGCGACGGGTGCATACATATTCCTCAGGGAACCTTGGTCCGCAAGAGTGCCTCAAAATG 5687
QY 68 GlnAlaAlaValAspValHisAsnAlaGln-----IleThrAlaLeuIleGluThrArg 85
Db 5688 CAGGCGCGCTGTACGCTGCAGGGCTGCAGATTCCGGTATCCGGCCGACCAAGCTCGTACT 5747
QY 86 AlaSerArgLeuHisPheGluGlu---ThrProAlaThrIleAlaAspThrPheAla 104
Db 5748 TCGGCGCGGCTGTGCTGTCGGGTGCGATGTGCGCCAGCCCTCCAGCCTGACACACGCGG 5807
QY 105 LysAlaGluLysLeuAspArgLeuAlaThrThrThrSerGlyAlaLeuArgAlaThrPro 124
Db 5808 GT-CAGCATCAGCGTGAACGAAACAGTTTCATCGGACGCGCGGCTGACCCACACACCC 5866
QY 125 PheAlaMetAlaSerLeuLeuGlnTyrMetGlnProAlaIleAsnLysGlyAspTrpLeu 144
Db 5867 TACATTGTGGCGCGCTTCTGTGATATCAACCCCATCTTTCATCAG-----TGCCA 5920
QY 145 ProAlaProLeuLysProLeuThrProLeuIleSerGlyAlaLeuSerGlyAlaMetAsp 164
Db 5921 CCGGCTGAATTTCAG---CTTCCCTCAAGCTCGAAGG---CTGCACGAGACCCAGCA 5974
QY 165 GlnVal-----GlyThrLysMetMetAspArgAlaThrGlyAspLeuHisTyrLeuSer 182
Db 5975 CGGGTGTCTTGGGCCAGCGGTGCCGAACATTCGAGCTGCAGAT---CACCAACATCCC 6031
QY 183 AlaSerProAspArgLeuHisAspAlaMetAlaAlaSerVal----- 196
Db 6032 GCTTCCCCCGCATCGCCAGCTGTGCTGCTGCCCCAGCGCTTCTTCGACAAAGACGTAA 6091
QY 197 -----LysArgHisSerProSer-LeuAlaArg-----GlnValLeuAsp-- 209
Db 6092 GCAGAGATCGTCATTCGTTCTGTTCTCGCACAGACATTTGATCCGGAAGTGTGAAGC 6151
QY 210 -ThrGlyValAlaValGlnThrTyrSerAlaArgAsnAlaValArgThrValLeuAlaPr 229
Db 6152 GACGGGCATCTCGCATCGTGTGTCGGCCAGCAGACCGAATTACCGTGCCTACCTTCCC 6211
QY 229 oAlaLeuAlaSerArgProAlaValGlnGlyAlaValAspLeuGlyValSer-----Me 247
Db 6212 GGTCTTCTCCACCATTCGCAACAGGGCAACGCGGTGTGTGCGCGTGTGACGAGACT 6271
QY 247 tAlaGlyGlyLeuAlaAlaAsnAlaGlyPheGlyAsnArgLeuLeuSerValGlnSerAr 267
Db 6272 GCCTTCCGCCCTCGGGCGCCAGCGGTCAGTGGCCCAACGCTTATGGAAGTGGCAATCC 6331
QY 267 gasPHisGlnArgGlyGlyAlaLeuValLeuGlyLeuLysAspLysGluProLysAlaGl 287
Db 6332 ATCCGACCCCAACGGCAGCATCTGCTGTACCCGGCGCGCAACCGTGAT----- 6380
QY 287 nLeuSerGlnGluAsnAspTrpLeuGluAlaTyrLysAlaIleLysSerAlaSerTyrSe 307
Db 6381 -----GAAGTCATCACCGGAGCAAGGCAATCGGTTTGG 6415
QY 307 rGlyAlaAlaLeuAsnAlaGlyLysArgMetAlaGlyLeuProLeuAsp----- 323
Db 6416 TTCGAGCACCTGCGCAGACCAACCGCATGACGTGGCGCGCATCGAGTGGGCGCCG 6475
QY 324 -MetAlaThrAspAlaMetGlyAlaValArgSer-----Le 335
Db 6476 CGTGGCAATGACGCGCCCTCTTCATTCGACCAACCGCCCGGTCGCGCTGGGGAAC 6535

QY 335 uValSerAlaSerSerLeuThrGlnAsnGlyLeuAla 347
Db 6536 GGTGCCAGACAGCCCGCTGCAGCGCTGAAGGTACGCC 6572

RESULT 15

US-09-147-236-10
; Sequence 10, Application US/09147236A
; Patent No. 6316251

GENERAL INFORMATION:

; APPLICANT: TONOUCHI, Naoto
; APPLICANT: TSUCHIDA, Takayasu
; APPLICANT: YOSHINAGA, Fumihito
; APPLICANT: TAHARA, Naoki

; APPLICANT: HAYASHI, Takahisa

; TITLE OF INVENTION: NOVEL GENE, GROUP OF GENES, AND NOVEL BETA-GLUCOSIDASE

; FILE REFERENCE: 6537-011-0PCT

; CURRENT APPLICATION NUMBER: US/09/147, 236A

; EARLIER APPLICATION NUMBER: PCT/JP97/03633

; NUMBER OF SEQ ID NOS: 12

; SOFTWARE: Patentln Ver. 2.1

; SEQ ID NO 10

; LENGTH: 16836

; TYPE: DNA

; ORGANISM: Acetobacter xylinum

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (1891)..(2922)

; FEATURE:

; OTHER INFORMATION: Nucleotide sequence is the same as SEQ ID NO:1

; OTHER INFORMATION: t

US-09-147-236-10

Alignment Scores:

Pred. No.: 0.0852

Score: 128.00

Percent Similarity: 39.95%

Best Local Similarity: 25.20%

Query Match: 5.31%

DB: 4

US-09-825-414-7 (1-486) x US-09-147-236-10 (1-16836)

QY 10 ProProValThrAlaThrAspSerPheArgThrAlaSerAspAlaSerLeuAlaSerSer 29

Db 5535 CCGCCCGCCACAAACC-----CCGGCGCGCGATGCTCAGCAGCCAGCGCG 5579

QY 30 SerValArgSerValSerSerAspGlnGlnArgGluIleAsnAlaIleAla-----Asp 47

Db 5580 GCTGATGCGGTGTGGACAATGCCGAG-----AACGCCATCGCCGGGTCTGAC 5627

QY 48 TyrLeuThrAspHisValPheAlaAlaHisLysLeuProProAlaAspSerAlaAspGly 67

Db 5628 GTGGCGACGGTGCATACATATTCCCTCAGGGAATTGGTGCAGAGTGCCTCAAAATG 5687

QY 68 GlnAlaAlaValAspValHisAsnAlaGln-----IleThrAlaLeuIleGluThrArg 85

Db 5688 CAGGGCGCTGCTACGCTGCAGGGCTGCAGTTCGGTATTCGGCGCCGACGACTCGTGACT 5747

QY 86 AlaSerArgLeuHisPheGluGlyGlu--ThrProAlaThrIleAlaAspThrPheAla 104

Db 5748 TCGGCGCGGCTTGTGCTGTCGGGTGCAGATGTCGCCAGCCTCCAGCCTGACACCGCGG 5807

QY 105 LysAlaGluLysLeuAspArgLeuAlaThrThrThrSerGlyAlaLeuArgAlaThrPro 124

Db 5808 GT-CACGATCACGCTGAACAGACAGTTCATCGGACGCTGCGGCTGACCCACACACACC 5866

QY 125 PheAlaMetAlaSerLeuLeuGlnTyrMetGlnProAlaIleAsnLysGlyAspTrpLeu 144

Db 5867 TACATTTGGCGCGCTTTCGTTTGATATCAACCCCATCTTCTCATCAG-----TGGCAA 5920

QY 145 ProAlaProLeuLysProLeuThrProLeuIleSerGlyAlaLeuSerGlyAlaMetAsp 164

Db 5921 CCGGCTGAATTTCAG---CTTCGCTTCAAGCTCGAAGGG---CTGCACGGAACCCAGCAA 5974

QY 165 GlnVal-----GlyThrLysMetMetAspArgAlaThrGlyAspLeuHisTyrLeuSer 182

Db 5975 CCGGTGTCTGCGCCAGCGGTGTCGGAACATTCCGAGCTGCAGAT---CACCAACATCCC 6031

QY 183 AlaSerProAspArgLeuHisAspAlaMetAlaAlaSerVal----- 196

Db 6032 GCTTCCCCCGCATGCCAGCTGTCCGCTGCCCCCAGCCGTTCTTCGACAGAAGCTAA 6091

QY 197 -----LysArgHisSerProSer-LeuAlaArg-----GlnValLeuAsp-- 209

Db 6092 GCAGAAGATCGTCAATTCCGTTCTGTCGACAGACATTGTGATCCCGAAGTGTGAAGGC 6151

QY 210 -ThrGlyValAlaValGlnThrTyrSerAlaArgAsnAlaValArgThrValLeuAlaPr 229

Db 6152 GACGGCATCCGTGCATCGTGTTCGCCACAGACCGATTACCGTGGCGTCACTTCCC 6211

QY 229 oAlaLeuAlaSerArgProAlaValGlnGlyAlaValAspLeuGlyValSer-----Me 247

Db 6212 GGTCTTCTCCACCAATTCCGCAACAGCGGCAACGCCGTGTGTCCGCGCTGACGAGCT 6271

QY 247 tAlaGlyGlyLeuAlaAlaAsnAlaGlyPheGlyAsnArgLeuLeuSerValGlnSerAr 267

Db 6272 GCCTTCGCCCGCTCGCGGCCAGCGGTCAGTGGCCCCACGCTTATGAGTGGCCATCC 6331

QY 267 gasPHisGlnArgGlyGlyAlaLeuValLeuGlyLeuLysAspLysGluProLysAlaGl 287

Db 6332 ATCCGACCCCAACGGCAGATCCTGCTCTGTAACCGGGCGGACCGGTGAT----- 6380

QY 287 nLeuSerGluGlnAsnAspTrpLeuGluAlaTyrLysAlaIleLysSerAlaSerTyrSe 307

Db 6381 -----GAAGTCATCACCGCGAGCAAGGCGATCGGTTTGG 6415

QY 307 rGlyAlaAlaLeuAsnAlaGlyLysArgMetAlaGlyLeuProLeuAsp----- 323

Db 6416 TTCGAGCACCTGCGCGACAGCAACCGCATGAGCTGGCGCGATCGAGGTGCGGCGCG 6475

QY 324 -MetAlaThrAspAlaMetGlyAlaValArgSer-----Le 335

Db 6476 CGTGGCAATGACGCGCCCTCTTCATTTCCGACCAACGCCCGCGCTCGGCGGACT 6535

QY 335 uValSerAlaSerSerLeuThrGlnAsnGlyLeuAla 347

Db 6536 GGTGCCAGACAGCGCCCTGCAGGCTGAAGGTTCACGCC 6572

Search completed: January 31, 2003, 08:53:21

Job time : 5780.44 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: January 31, 2003, 04:32:35 ; Search time 76.9209 Seconds
(without alignments)
2838.537 Million cell updates/sec

Title: US-09-825-414-7

Perfect score: 2412
Sequence: 1 MHINRRVQPPVTATDSFRT.....IEEGTAASPSLIFRPMRS 486

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 396772 seqs, 224632407 residues

Total number of hits satisfying chosen parameters: 793544

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ .p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO.spool/US09825414/runat_28012003_161742_26139/app_query.fasta_1.1294
-DB=Published Applications_NA -QFMT=fastap -SUFFIX=p2n.rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=45 -DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=plto -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USER=US09825414@cgn2_1_1_46@runat_28012003_161742_26139
-NCPU=6 -ICPU=3 -NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published_Applications_NA:*

1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2412	100.0	1461	10	US-09-825-414-6
2	2412	100.0	30365	10	US-09-825-414-1
3	1869.5	77.5	1464	10	US-09-825-414-65
4	148.5	6.2	4645	9	US-09-971-536-27

5	143	5.9	1801	10	US-09-287-849-21	Sequence 21, Appl
6	133.5	5.5	2287	10	US-09-287-849-1	Sequence 1, Appl
7	133.5	5.5	5036	10	US-09-918-951-2	Sequence 2, Appl
8	131.5	5.5	6258	10	US-09-815-242-4584	Sequence 4584, Ap
9	131.5	5.5	17388	10	US-09-815-242-8512	Sequence 8512, Ap
10	129.5	5.4	7035	10	US-09-815-242-8615	Sequence 8615, Ap
11	129	5.3	7302	10	US-09-815-242-4780	Sequence 4780, Ap
12	129	5.3	18846	10	US-09-815-242-8898	Sequence 8898, Ap
13	127.5	5.3	1797	10	US-09-287-849-25	Sequence 25, Appl
14	127.5	5.3	11220	9	US-09-860-846-32	Sequence 32, Appl
15	127.5	5.3	11220	10	US-09-861-289-32	Sequence 32, Appl
16	127.5	5.3	15872	9	US-09-860-846-1	Sequence 1, Appl
17	127.5	5.3	15872	10	US-09-861-289-1	Sequence 1, Appl
18	127.5	5.3	36778	9	US-09-860-846-5	Sequence 5, Appl
19	127.5	5.3	36778	10	US-09-861-289-5	Sequence 5, Appl
20	127	5.3	1248	10	US-09-974-300-902	Sequence 902, App
21	127	5.3	3300	10	US-09-379-931-6	Sequence 6, Appl
22	127	5.3	9542	9	US-09-813-214A-8	Sequence 8, Appl
23	126.5	5.2	32768	10	US-09-070-927A-399	Sequence 399, App
24	125	5.2	2141	10	US-09-841-786-11	Sequence 11, Appl
25	125	5.2	9726	10	US-09-841-786-8	Sequence 8, Appl
26	125	5.2	11130	10	US-09-841-786-15	Sequence 15, Appl
27	124.5	5.2	7995	10	US-09-864-864-333	Sequence 333, App
28	123.5	5.1	6177	9	US-10-124-800-3	Sequence 3, Appl
29	123.5	5.1	8078	10	US-09-791-942-3	Sequence 3, Appl
30	120.5	5.0	7639	9	US-09-971-536-28	Sequence 28, Appl
31	120	5.0	8730	9	US-10-124-800-1	Sequence 1, Appl
32	119.5	5.0	2155	10	US-09-960-428-13	Sequence 13, Appl
33	119.5	5.0	4645	9	US-09-971-536-27	Sequence 27, Appl
34	118.5	4.9	3954	10	US-09-815-242-4063	Sequence 4063, Ap
35	118	4.9	7100	10	US-09-932-183A-1	Sequence 1, Appl
36	117.5	4.9	2331	10	US-09-815-242-4037	Sequence 4037, Ap
37	117	4.9	9319	10	US-09-956-004-85	Sequence 85, Appl
38	116.5	4.8	1380	10	US-09-815-242-7856	Sequence 7856, Ap
39	116.5	4.8	9210	9	US-09-712-363-100	Sequence 100, App
40	116	4.8	2523	10	US-09-476-242-15	Sequence 15, Appl
41	116	4.8	7434	10	US-09-815-242-4761	Sequence 4761, Ap
42	116	4.8	7437	10	US-09-815-242-8869	Sequence 8869, Ap
43	116	4.8	8979	9	US-09-738-626-2739	Sequence 2739, Ap
44	116	4.8	68750	9	US-10-014-717-1	Sequence 1, Appl
45	115	4.8	1029	10	US-09-815-242-6946	Sequence 6946, Ap

ALIGNMENTS

RESULT 1
US-09-825-414-6
; Sequence 6, Application US/09825414
; Patent No. US20020083489A1
; GENERAL INFORMATION:
; APPLICANT: Collier, Alan
; APPLICANT: Alfano, James R.
; TITLE OF INVENTION: DNA MOLECULES AND POLYPEPTIDES OF PSEUDOMONAS SYRINGAE
; TITLE OF INVENTION: HRP PATHOGENICITY ISLAND AND THEIR USES
; FILE REFERENCE: 19603/3243
; CURRENT FILING DATE: 2001-04-03
; PRIOR APPLICATION NUMBER: 60/194,160
; PRIOR FILING DATE: 2000-04-03
; PRIOR APPLICATION NUMBER: 60/224,604
; PRIOR FILING DATE: 2000-08-11
; PRIOR APPLICATION NUMBER: 60/249,548
; NUMBER OF SEQ ID NOS: 91
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 6
; LENGTH: 1461
; TYPE: DNA
; ORGANISM: Pseudomonas syringae
US-09-825-414-6

Alignment Scores:

Pred. No.:	1.03e-224	Length:	1461
Score:	2412.00	Matches:	486
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0
DB:	10	Gaps:	0
US-09-825-414-7 (1-486) x US-09-825-414-6 (1-1461)			
QY	1	MethisIleAsnArgValGInGInProProValThrAlaThrAspSerPheArgThr	20
Db	1	ATGCACATCAACCGACGCGCTCCACACACCGCCTGTGACTGCGACGGATAGCTTTCGGACA	60
QY	21	AlaSerAspAlaSerLeuAlaSerSerSerValArgSerValSerSerAspGInGInArg	40
Db	61	GCGTCCGACGCGTCTTGCTTCCTCCAGCTCTGTGCGATCTGTACAGCTCCGATCAGCAACGC	120
QY	41	GIuIleAsnAlaIleAlaAspTyrLeuThrAspHisValPheAlaAlaHisLysLeuPro	60
Db	121	GAGATAAATGCATGCGCATTAACCTGACAGATCATGTTCGCTGCCATAAACTGCCG	180
QY	61	ProAlaAspSerAlaAspGlyGInAlaAlaValAlaAspValHisAsnAlaGInIleThrAla	80
Db	181	CCGGCCGATTCGGCTGATGGCCAAGCTGCAGTTGACGTACACAATGCCAGATCAGTCCG	240
QY	81	LeuIleGluThrArgAlaSerArgLeuHisPheGluGlyGluThrProAlaThrIleAla	100
Db	241	CTGATCGAGACGCGCGCCAGCCGCTGCACCTCGAAGGGGAAACCCCGGCAACCATCGCC	300
QY	101	AspThrPheAlaLysAlaGluLysLeuAspArgLeuAlaThrThrThrSerGlyAlaLeu	120
Db	301	GACACCTTCGCCAAGCGCGGAAAGCTCGACCGATTGGCGACGACTACATCAGCGCGCTTG	360
QY	121	ArgAlaThrProPheAlaMetAlaSerLeuLeuGInTyrMetGInProAlaIleAsnLys	140
Db	361	CGGGCGACGCGCTTGGCATGGCCCTCGTTGCTTCACTACATGACCGCTGCATCAACAAG	420
QY	141	GlyAspTrpLeuProAlaProLeuLysProLeuThrProLeuIleSerGlyAlaLeuSer	160
Db	421	GCGCATTTGGCTGCGCGCTCCCTCAAAACCGCTGACCCGCTCATTTCCGAGCGCTGCG	480
QY	161	GlyAlaMetAspGlnValGlyThrLysMetMetAspArgAlaThrGlyAspLeuHisTyr	180
Db	481	GGCGCCATGGACCAAGTGGGCAACAGATGATGACCGCGCGAGCGGTGATCTGCATTAC	540
QY	181	LeuSerAlaSerProAspArgLeuHisAspAlaMetAlaAlaSerValLysArgHisSer	200
Db	541	CTGAGCGCGCTCGCCGACAGGCTCCACGATGCGATGGCGCTTCGGTGAAGCGCCACTCG	600
QY	201	ProSerLeuAlaArgGlnValLeuAspThrGlyValAlaValGInThrTyrSerAlaArg	220
Db	601	CCAAGCCTTGGCTCGACAGGTTCTGGACACGGGGGTTGCGGTTTCAGACGTACTCGCGCGC	660
QY	221	AsnAlaValArgThrValLeuAlaProAlaLeuAlaSerArgProAlaValGInGlyAla	240
Db	661	AACGCCGTACGTACCGTATGTGCTCCGGCACTGGCGTCAGACCCGCGTGACGGTGCT	720
QY	241	ValAspLeuGlyValSerMetAlaGlyGlyLeuAlaAlaAsnAlaGlyPheGlyAsnArg	260
Db	721	GTGACCTTGGTGTATCGATGGCGGGGTGCTGTGCGTCCAAACGACGCTTTGGCAACCGC	780
QY	261	LeuLeuSerValGInSerArgAspHisGInArgGlyAlaLeuValLeuGlyLeuLys	280
Db	781	CTGCTCAGTGTGACGTGGGTGATCACACGCGTGGCGGTGCATTAGTGTGCTGTTGAAG	840
QY	281	AspLysGluProLysAlaGInLeuSerGluGluAsnAspTrpLeuGluAlaTyrLysAla	300
Db	841	GATAAAGAGCCCAAGGCTCACTGAGCGAAGAAAACGACTGGCTCGAGGCTTATTAAGCA	900
QY	301	IleLysSerAlaSerTyrSerGlyAlaAlaLeuAsnAlaGlyLysArgMetAlaGlyLeu	320
Db	901	ATCAAAATCGCCAGCTACTCGGGTGCGCGGCTCAACGCTGGCAAGCGGATGGCGGCTCG	960

QY	321	ProLeuAspMetAlaThrAspAlaMetGlyAlaValArgSerLeuValSerAlaSerSer	340
Db	961	CCACTGATATGGCGACCGACCAATGGGTGCGGTAGAAAGCCTGGTGTACAGCGTCCAGC	1020
QY	341	LeuThrGlnAsnGlyLeuAlaLeuAlaGlyGlyPheAlaGlyValGlyLysLeuGInGlu	360
Db	1021	CTGACCCAAAACGGTCTGGCCCTGGCGGGGTGGCTTTCAGGGGTAGGCAAGTTCAGGAG	1080
QY	361	MetAlaThrLysAsnIleThrAspProAlaThrLysAlaAlaValSerGInLeuThrAsn	380
Db	1081	ATGGCGACGAAAAATATATACCGACCGCGGACCAAGCCGCGGTCAGTTCAGTTCACCAAC	1140
QY	381	LeuAlaGlySerAlaAlaValPheAlaGlyTyrThrThrAlaAlaLeuThrThrAspPro	400
Db	1141	CTGGCAGGTTCCGACCGCGTTTTCGACAGGCTGGACCAAGCCGCGGTGACACCGATCCC	1200
QY	401	AlaValLysLysAlaGluSerPheIleGlnAspThrValLysSerThrAlaSerSerThr	420
Db	1201	GCGGTGAATAAACCGAGTCTGTTCATACAGACACGGTGAATCGACTGCATCCAGTACC	1260
QY	421	ThrGlyTyrValAlaAspGInThrValLysLeuAlaLysThrValLysAspMetGlyGly	440
Db	1261	ACAGGCTACGTAGCCGACCAAGCCGTCAAACCTGGCGAAGACCGTCAAGACATGGCGGG	1320
QY	441	GluAlaIleThrHisThrGlyAlaSerLeuArgAsnThrValAsnLeuArgGInArg	460
Db	1321	GAGCGATCACCATACCGCGCCAGCTTGCGCAATACGCTCAATACCTGCGTCAACGC	1380
QY	461	ProAlaArgGluAlaAspIleGluGluGlyGlyThrAlaAlaSerProSerGluIlePro	480
Db	1381	CCGGCTCGTGAAGCTGATATAGAAGAGGGGGGACAGCGGCGCTTCTCCAAGTGAATACCG	1440
QY	481	PheArgProMetArgSer	486
Db	1441	TTTCGGCCTATGCGGCTCG	1458

RESULT 2

US-09-825-414-1/c

; Sequence 1, Application US/09825414

; Patent No. US20020083489A1

; GENERAL INFORMATION:

; APPLICANT: Colmer, Alan

; APPLICANT: Alfano, James R.

; APPLICANT: Charkowski, Amy O.

; TITLE OF INVENTION: DNA MOLECULES AND POLYPEPTIDES OF PSEUDOMONAS SYRINGAE

; FILE REFERENCE: 19603/3243

; CURRENT APPLICATION NUMBER: US/09/825,414

; PRIOR FILING DATE: 2001-04-03

; PRIOR APPLICATION NUMBER: 60/194,160

; PRIOR FILING DATE: 2000-04-03

; PRIOR APPLICATION NUMBER: 60/224,604

; PRIOR FILING DATE: 2000-08-11

; PRIOR APPLICATION NUMBER: 60/249,548

; NUMBER OF SEQ ID NOS: 91

; SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 1

; LENGTH: 30365

; TYPE: DNA

; ORGANISM: Pseudomonas syringae

; FEATURE:

; NAME/KEY: unsure

; LOCATION: (29734)

; OTHER INFORMATION: n at any position is undefined

; US-09-825-414-1

Alignment Scores:

Pred. No.:	6.05e-223	Length:	30365
Score:	2412.00	Matches:	486
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	100.00%	Indels:	0

DB: 10 Gaps: 0
US-09-825-414-7 (1-486) x US-09-825-414-1 (1-30365)
QY 1 MethIstIleasnArgArgValGInGInProProValThrAlaThrAspSerPheArgThr 20
DB 23984 ATGCACATCAACCGACGCGCTCCACAAACCGCCTGTGACTCGCAGCGATAGCTTTCGACA 23925
QY 21 AlaserAspAlaserSerleuAlaserSerSerValArgSerValSerSeraspGInGInArg 40
DB 23924 GCGTCGACGCGCTCTTGGCTCCAGCTCTGTGCGATCTGTACAGTCCGATCAGCAACGC 23865
QY 41 GluIleAsnAlaIleAlaAspTyrLeuThrAspHisValPheAlaAlaHisLysLeuPro 60
DB 23864 GAGATAAATGCCGATTGCCGATTACTGACAGATCATGTGTTGCTGCGCATPAAACTGCGC 23805
QY 61 ProAlaaspSerAlaaspGlyGInAlaAlaValaspValHisasnAlaGInIleThrAla 80
DB 23804 CCGGCGGATTGCGGTGATGGCAAGCTGCAGTTGACGTACACAAATGCCAGATCACTGCC 23745
QY 81 LeuIleGluThrArgAlaSerArgLeuHisPheGluGlyGluThrProAlaThrIleAla 100
DB 23744 CTGATCGAGACGCGCGCCAGCGCGCTGCACCTTCGAAGGGAAACCCGCGCAACCATCGCC 23685
QY 101 AspThrPheAlaLysAlaGluLysLeuAspArgLeuAlaThrThrSerGlyAlaLeu 120
DB 23684 GACACCTTCGCCCAAGCGGAAAGCTGCACCGATTGGCGACGACTACATCAGCGCGCTTG 23625
QY 121 ArgAlaThrProPheAlaMetAlaSerLeuLeuGInTyrMetGInProAlaIleAsnLys 140
DB 23624 CCGGCGACGCCCTTTGCCATGGCCCTGCTTCTCAGTACATGCAGCCTGCGATCAACAAG 23565
QY 141 GlyAspTyrLeuProAlaProLeuLysProLeuThrProLeuIleSerGlyAlaLeuSer 160
DB 23564 GCGGATTGGCTGCCGCTCCGCTCAAAACGCTGACCCCGCTCATTTCCGGAGCGCTGTCG 23505
QY 161 GlyAlaMetaspGInValGlyThrLysMetMetaspArgAlaThrGlyAspLeuHisTyr 180
DB 23504 GCGCGCATGCACAGGTGGGCACCAAGATGATGACCGCGCGAGCGGTGATCTGCATTAC 23445
QY 181 LeuSerAlaserProaspArgLeuHisaspAlaMetAlaAlaserValLysArgHisSer 200
DB 23444 CTGAGCGGCTCGCCGAGACAGCTCCAGATGCGCTCGCGTTGGAAGCGCACACTCG 23385
QY 201 ProSerLeuAlaArgGInValLeuaspThrGlyValAlaValGInThrTyrSerAlaArg 220
DB 23384 CCAAGCCTTGCTCGACAGGTTCTGCACACGGGGGTTGCGGTTCAAGCTACTCGCGCGC 23325
QY 221 AsnAlaValArgThrValLeuAlaProAlaLeuAlaSerArgProAlaValGInGlyAla 240
DB 23324 AACGCCGTACGTACCGTATTGGCTCCGGCACTGGCGTCCAGACCCCGCTGCAGGGTGT 23265
QY 241 ValaspLeuGlyValSerMetAlaGlyGlyLeuAlaAlaasnAlaGlyPheGlyAsnArg 260
DB 23264 GTGACCTTGGTGTATCGATGGCGGTGCTTGGCTGCCAACGCAAGCTTTGGCAACCGC 23205
QY 261 LeuLeuSerValGInSerArgaspHisGInArgGlyGlyAlaLeuValLeuGlyLeuLys 280
DB 23204 CTGCTCAGTGTGACAGTCGCGTATCACACGCGTGGCGGTGCATTAGTCTCGGTTGAAG 23145
QY 281 AspLysGluProLysAlaGInLeuSerGluGluAsnAspTyrPleuGluAlaTyrLysAla 300
DB 23144 GATAAAGAGCCCAAGGCTCACTGACGGAAGAAACGACTGGCTTCGAGGCTTATAAGCA 23085
QY 301 IleLysSerAlaSerTyrSerGlyAlaAlaLeuAsnAlaGlyLysArgMetAlaGlyLeu 320
DB 23084 ATCAAAATCGGCCAGCTACTCGGGTGGCGGCTCAACGCTGGCAAGCGGATGGCGGTTCTG 23025
QY 321 ProLeuaspMetAlaThraspAlaMetGlyAlaValArgSerLeuValSerAlaSerSer 340
DB 23024 CCACTGGATATGGCGACCGACGCAATGGGTGCGGTAAAGAAGCCTGGTGTCAAGCTCCAGC 22965
QY 341 LeuThrGInasnGlyLeuAlaLeuAlaGlyGlyPheAlaGlyValGlyLysLeuGInGlu 360

DB 22964 CTGACCCAAACGGTCTGGCCCTGGCGGGTGGCTTTGCAGGGGTAGCCAAAGTTGCAGAG 22905
QY 361 MetAlaThrLysAsnIleThraspProAlaThrLysAlaAlaValSerGInLeuThrAsn 380
DB 22904 ATGGCGACGAAATATACCGACCGCGGACCAAGCGCGGTGATGATGACCAAC 22845
QY 381 LeuAlaGlySerAlaAlaValPheAlaGlyTyrThrThrAlaAlaLeuThrThrAspPro 400
DB 22844 CTGGCAGGTTGCGACGCGGTTTTCGAGCGCTGGACACAGCGCGCGCTGACAAACGATCCC 22785
QY 401 AlavalLysLysAlaGluSerPheIleGInAspThrValLysSerThrAlaSerSerThr 420
DB 22784 GCGGTGAATAAACCGAGTCTTCATACAGGACACGGTGAATCGACTGCATCCAGTACC 22725
QY 421 ThrGlyTyrValAlaaspGInThrValLysLeuAlaLysThrValLysaspMetGlyGly 440
DB 22724 ACAGGCTACGTAGCCGACACGACCGTCAAACTGGCGAAGACCGTCAAGACATGGCGGG 22665
QY 441 GluAlaIleThrHisThrGlyAlaSerLeuArgAsnThrValAsnAsnLeuArgGInArg 460
DB 22664 GAGGCGATACCCATACCGGCGCCAGCTTGGCAATACGGTCAATTAACCTGCGTCAACGC 22605
QY 461 ProAlaArgGluAlaaspIleGluGlyGlyThrAlaAlaSerProSerGluIlePro 480
DB 22604 CCGGCTCGTGAAGCTGATATAGAAGAGGGGGCACGCGCGCTTCCTCAAGTGAATACCG 22545
QY 481 PheArgProMetArgSer 486
DB 22544 TTTCCGCCCTATGCGGTGCG 22527
RESULT 3
US-09-825-414-65
; Sequence 65, Application US/09825414
; Patent No. US20020083489A1
; GENERAL INFORMATION:
; APPLICANT: Colimer, Alan
; APPLICANT: Alfano, James R.
; APPLICANT: Charkowski, Amy O.
; TITLE OF INVENTION: DNA MOLECULES AND POLYPEPTIDES OF PSEUDOMONAS SYRINGAE
; FILE REFERENCE: 19603/3243
; CURRENT APPLICATION NUMBER: US/09/825,414
; CURRENT FILING DATE: 2001-04-03
; PRIOR APPLICATION NUMBER: 60/194,160
; PRIOR FILING DATE: 2000-04-03
; PRIOR APPLICATION NUMBER: 60/224,604
; PRIOR FILING DATE: 2000-08-11
; PRIOR APPLICATION NUMBER: 60/249,548
; PRIOR FILING DATE: 2000-11-17
; NUMBER OF SEQ ID NOS: 91
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 65
; LENGTH: 1464
; TYPE: DNA
; ORGANISM: Pseudomonas syringae pv. tomato
; US-09-825-414-65
Alignment Scores:
Pred. No.: 3,54e-172 Length: 1464
Score: 1869.50 Matches: 381
Percent Similarity: 85.19% Conservative: 33
Best Local Similarity: 78.40% Mismatches: 71
Query Match: 77.51% Indels: 1
DB: 10 Gaps: 1
US-09-825-414-7 (1-486) x US-09-825-414-65 (1-1464)
QY 1 MethIstIleasnArgArgValGInGInProProValThrAlaThrAspSerPheArgThr 20
DB 1 ATGCACATCAACCAATCCGCCCAACAACCGCCTGGCGTTGCAATGACAGATTTCGACACA 60
QY 21 AlaserAspAlaserSerleuAlaserSerSerValArgSerValSerSeraspGInGInArg 40

Db 61 GCTCCGACGCGTCCCTGCTTCGAGTCTCTGCGCGTCTGTCAGCACTACCTCGTGGCCG 120
QY 41 GluIleAsnAlaIleAlaAspTyrLeuThrAspHisValPheAlaAlaHisLysLeuPro 60
Db 121 GATCTACAAGCTATTACCGATTAATCTGAAACATCACGTGTCCGTCGACAGCTTTTCG 180
QY 61 ProAlaAspSerAlaAspGlyGlnAlaAlaValAspValHisAsnAlaGlnIleThrAla 80
Db 181 GTAATAGGCTCACCAGATGAGCGTGATGCCGCTCTTGCAACACAGAGAGATGATGCCG 240
QY 81 LeuIleGluThrArgAlaSerArgLeuHisPheGluGlyLeuThrProAlaThrIleAla 100
Db 241 TTGGTAGAGACACGCCCAACCGCTTACTCCGAAGGGAGAGACCCCGCAACCATCGCC 300
QY 101 AspThrPheAlaLysAlaGluLysLeuAspArgLeuAlaThrThrSerGlyAlaLeu 120
Db 301 GAAACATTCGCCAAGCGGAAAGTTTCGACCGTTTGGCAGACCGCATCAAGTGTCTTT 360
QY 121 ArgAlaThrProPheAlaMetAlaSerLeuLeuGlnTyrMetGlnProAlaIleAsnLys 140
Db 361 GAGAACACGCGCATTTGCCGTGCGCTCGGTGCTTCAGTACATGCAGCCTGCGATCAACA 420
QY 141 GlyAspTyrPleuProAlaProLeuLysProLeuThrProLeuIleSerGlyAlaLeuSer 160
Db 421 GCGGATTTGGCTAGCAACGCGCTCAAGCCGCTGACCCCGCTCATTTCCGGAGCGCTGCG 480
QY 161 GlyAlaMetAspGlnValGlyThrLysMetMetAspArgAlaThrGlyAspLeuHisTyr 180
Db 481 GGAGCCATGGACCAAGGTGGGACCAAAATGATGATCGTGGGAGGGGTGATCTGCATTTAC 540
QY 181 LeuSerAlaSerProAspArgLeuHisAspAlaMetAlaAlaSerValLysArgHisSer 200
Db 541 CTGAGCACTTCGCCGGACAACTTGATGATGCGATGGCCGTATCGGTGAAGCGCCACTCG 600
QY 201 ProSerLeuAlaArgGlnValLeuAspThrGlyValAlaValGlnThrTyrSerAlaArg 220
Db 601 CCTCGCCTTGGTGCACAGGTGTGACATGGGATGCGATGCGACAGCGTTCTCGCGCTA 660
QY 221 AsnAlaValArgThrValLeuAlaProAlaLeuAlaSerArgProAlaValGlnGlyAla 240
Db 661 AATGTGTCGTACCGTATGCTCCAGCACTAGCGTCCAGACCGCTGCGTGCAGGGTGTCT 720
QY 241 ValAspLeuGlyValSerMetAlaGlyGlyLeuAlaAlaAsnAlaGlyPheGlyAsnArg 260
Db 721 GTGATTTTGGCGGTATCTACGGGGGTGGCTTGGTGGCAATGCAGGCTTTGGCGCACCG 780
QY 261 LeuLeuSerValGlnSerArgAspHisGlnArgGlyGlyAlaLeuValLeuGlyLeuLys 280
Db 781 ATGCTCAGTGTGCAATCGCCGCATCAACTGCGTGGGGGGGCATTCTGACTTGGCATGAAA 840
QY 281 AspLysGluProLysAlaGlnLeuSerGluGlnAsnAspTyrPleuGluAlaTyrLysAla 300
Db 841 GATTAAGAGCCCAAGGCCCGCTGAGTGAAGAACTGATTGGCTTGATGCTTACAAAGCG 900
QY 301 IleLysSerAlaSerTyrSerGlyAlaAlaLeuAsnAlaGlyLysArgMetAlaGlyLeu 320
Db 901 ATCAAGTCGGCCAGCTACTACGTGCGCGCTCAATGCGGCAAGCGGATGGCCGCGCTG 960
QY 321 ProLeuAspMetAlaThrAspAlaMetGlyAlaValArgSerLeuValSerAlaSerSer 340
Db 961 CCACTGGACGTCGCGACCGAGCGGCTCAAGCGGCTGAGAGAGTCTGGTGGCCACACAGC 1020
QY 341 LeuThrGlnAsnGlyLeuAlaLeuAlaGlyGlyPheAlaGlyValGlyLysLeuGlnGlu 360
Db 1021 CTGACAAAAAATGGCCCTGGCCCTAGCCGGTGTACGCCGGGTAAGTTGACAGAAA 1080
QY 361 MetAlaThrLysAsnIleThrAspProAlaThrLysAlaAlaValSerGlnLeuThrAsn 380
Db 1081 ATGGCAGCAAAAAATATCACTGATTGCGCGACCAAGGCTGCGTTAGTCAGCTGAGCAAC 1140
QY 381 LeuAlaGlySerAlaAlaValPheAlaGlyTyrThrThrAlaAlaLeuThrThrAspPro 400

Db 1141 CTGGTGGTTCGGTAGCGGTTTTCGACAGCGCTGACCAACCGCTGGACTGGCGACTGACCCT 1200
QY 401 AlValLysLysAlaGluSerPheIleGlnAspThrValLysSerThrAlaSerSerThr 420
Db 1201 GCGGTTAAGAAAGCCGATCGTTTATACAGGATAGGTGAATGCAGCGCATCTAGTACC 1260
QY 421 ThrGlyTyrValAlaAspGlnThrValLysLeuAlaLysThrValLysAspMetGlyGly 440
Db 1261 ACAAGCTATGTTGCCGACCAAGACCGCTCAAACTGGCGGAAAAAGTCAAGGACATGACGGG 1320
QY 441 GluAlaIleThrHisThrGlyAlaSerLeuArgAsnThrValAsnAsnLeuArgGlnArg 460
Db 1321 GAGCGGATCTCCAGCACCGGTGCCAGCTTACGCACTACTGCAATAACCTGCGTCATCGC 1380
QY 461 ProAlaArgGluAlaAspIleGluGluGlyGlyThrAlaAla---SerProSerGluIle 479
Db 1381 TCCGCTCCGGAAGCTGATATCGAAGAAGGTGGGATTTCCGCCGTTTCTCGAAGTGAACA 1440
QY 480 ProPheArgPrometArg 485
Db 1441 CCGTTTCAGCTCAGCGCT 1458

RESULT 4
US-09-971-536-27
; Sequence 27, Application US/09971536
; Patent No. US20020159976A1
; GENERAL INFORMATION:
; APPLICANT: Glenn, Matthew
; APPLICANT: Havukkala, Ilkka
; APPLICANT: Bloksberg, Leonard
; APPLICANT: Lubbers, Mark
; APPLICANT: Dekker, James
; APPLICANT: Christensson, Anna
; APPLICANT: Holland, Ross
; APPLICANT: O'Toole, Paul
; APPLICANT: Reid, Julian
; APPLICANT: Coolbear, Timothy
; TITLE OF INVENTION: Lactobacillus rhamnosus Polynucleotides, Polypeptides and Meth
; FILE REFERENCE: 1043c2
; CURRENT APPLICATION NUMBER: US/09/971,536
; CURRENT FILING DATE: 2001-10-02
; PRIOR APPLICATION NUMBER: U.S. No. US20020159976A1 09/634,238
; PRIOR FILING DATE: 2000-08-08
; PRIOR APPLICATION NUMBER: U.S. No. US20020159976A1 09/724,623
; PRIOR FILING DATE: 2000-11-28
; PRIOR APPLICATION NUMBER: PCT/NZ01/00160
; NUMBER OF SEQ ID NOS: 83
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 27
; LENGTH: 4645
; TYPE: DNA
; ORGANISM: Lactobacillus rhamnosus
US-09-971-536-27

Alignment Scores:
Pred. No.: 7.58e-05 Length: 4645
Score: 148.50 Matches: 126
Percent Similarity: 36.18% Conservative: 69
Best Local Similarity: 23.38% Mismatches: 233
Query Match: 6.16% Indels: 112
DB: 9 Gaps: 21

US-09-825-414-7 (1-486) x US-09-971-536-27 (1-4645)
QY 13 ThrAlaThrAspSerPheArgThrAla-----SerAspAlaSerLeuAlaSerSer--- 29
Db 2864 ACTGACACATCTCGTATGCTAGTGTGCTGCAAGTTCGAGTGCACAGTGAAGCCGGTTCATAT 2923
QY 30 -----SerValArgSerValSerSerAspGlnArgGluIleAsnAlaIleAlaAsp 47
Db 2924 GGTCAATCAGGACGAGGCTCCAGCGCCAGTGACGCTGTGGTCACTCCGCGAGTGCAGCCCAA 2983

QY 48 TyrLeuThrAspHisValPheAlaAlaHisLysLeuProPro-----Alaaspser 64
::: :::: ||||| |||
Db 2984 CATGCCAGCACCAGCTGCCAGTCCGATCCAGCTATCCGAAGATAGTGGATTTCAGTCA 3043
QY 65 AlaaspGlyGlnAlaAlaValaspValHisasnAlaGlnIleThrAlaLeuIleGlyThr 84
||| ||||| :::: |||
Db 3044 CTAGCCAGTCAAGCTGCCAAGCGCAGCAGCAAGGCAAGCAAGTAACGCGAGTCCGCAACC 3103
QY 85 ArgAlaSerArgLeuHisPheGluGlyGlyThrProAlaThr---IleAlaaspThrPhe 103
||| :::: ||| :::: ||| :::: |||
Db 3104 AGCGCCGCGCGCTGTTGTTTCAGTGTCCAGTGAAGTAACAGCGCAAGCGGCT 3163
QY 104 AlaLysAlaGlyLysLeuaspArgLeuAlaThrThr-----SerGly 118
||| ||| :::: ||| :::: |||
Db 3164 GCAAGTCCGATGTGTGGCAAGCAGTCCGCGCAGCAGCGCTAACAGTAATGCCAGTGC 3223
QY 119 AlaLeuArgAlaThrPro-----PheAlaMetAlaSer 129
||| ||||| ||| :::: ||| :::: |||
Db 3224 GCAGCCAGTCCGACCAAGCTGTGATAGCAAAAGCCGCGCAGAGATTCTCGAGTGCAGCG 3283
QY 130 -LeuLeuGlnIleTyrMetGlnProAlaIleasnLysGlyaspTrpLeuProAla----- 146
::: ||||| ||| :::: ||| ||||| |||
Db 3284 AGTCTGCAGCAAGCAGTGCAGGGTGCAGAG-----CAGTTGCCAGCGAAGCGCG 3337
QY 147 ----ProLeuLysProLeuThrProLeuIleSerGlyAlaLeuSerGlyAla-----Me 163
||| ||| :::: ||| :::: ||| ||| |||
Db 3338 AGTCCGCGGCGCATCGGATGCTCG-GTAGCTTCTAGTCCGCGCAGTCCGCGTGCAGGCTT 3396
QY 163 tAspGlnValGlyThrLysMetMetaspArgAlaThrGlyaspLeuHisTyrLeuSerAl 183
||| :::: ||| :::: |||
Db 3397 TGACAAAGCTGCCAGCGCTCGGGAAGCGCAGCTTCA-----AGTGC 3438
QY 183 aserProaspArgLeuHisaspAlaMetAlaAlaSerValLysArgHisSerProSerIle 203
||: ||: ||| ||| :::: |||
Db 3439 CGCGAGCGCGGCTGTAGTTCACGGCGCAGCTCAAGGCACACGAGGTGGCGCAAGCTCCAG 3498
QY 203 uAlaArgGlnValLeuaspThrGlyValAlaValGlnThrTyr----- 217
||| :::: |||
Db 3499 TGGCAGCGGAAGCGGTCAAGCATCAACCGCAACATCTGTGTATGTAGTGTGCAAGTTTC 3558
QY 218 -----SerAlaArgasnAlaVal 223
||| |||
Db 3559 GAGTGCACAGTGAAGCCGGTTCATATGCTCATCAGGCGAGGCTCCAGTGCAGTGAAGCGAC 3618
QY 223 1-----ArgThrValLeuAlaProAlaLeuAlaSerArgPr 235
::: ||| :::: ||||| |||
Db 3619 TGGCCATGCAAGTAGTGTACAAGTCAAGCAAGTCCGCGCATCCAGTGTGCTCCAGGTA 3678
QY 235 oAlaValGlnGlyAlaValaspLeuGlyValSerMetAlaGlyGlyLeuAlaAlaasnAl 255
::: ||: ||||| :::: |||
Db 3679 CCAAGTGTAGTGGGATCCAGTCAAGTGTAAGTATTCGCTCCAGTGCAGCAAGTACTGC 3738
QY 255 aglyPheGlyasnArgLeuLeuSerValGlnSerArgaspHisGlnArgGlyAlaIle 275
||| |||
Db 3739 ATCC-----AGTCCGCTAGTGGCGCACAAAGTGAGGCTTCGACGCG 3780
QY 275 uValLeuGlyLeuLysaspLysGluProLysAlaGlnLeuSerGluGluasnaspTrpIle 295
||| :::: ||| :::: |||
Db 3781 ATCGTCGGCTGCAAGTCAATGCTAGTGAACAAGCAAGTATTCCTTCAGTGAGGATGTTGT 3840
QY 295 uGluAlaTyrLysAlaIleLysSerAlaSerTyrSerGlyAlaAlaLeuasnAlaGlyLy 315
||| ||: ||| :::: ||| :::: |||
Db 3841 ATCAAGCAGTGTGCTGAGTGTGCGTCCAGCGCGCGCAGTCCGCGCATCCAGTGTGCCAA 3900
QY 315 s-----ArgMetAlaGlyLeuProLeuaspMetAlaThraspAlaMetGlyAl 331
||| ||| :::: |||
Db 3901 GGCTGTGTAACAGTAGTGTGCGGGTATATATCTCATGCGACCAAGTGACAGCTGCCAAGCAG 3960
QY 331 aValArgSerLeuValSerAlaSerSerLeuThrGlnasnGlyLeuAlaLeuAlaGlyGl 351
||: ||| ||| :::: |||
Db 3961 TGGTAAGAGCGGCTGAAGAGTCAAGCAAGCAGAGTGCAGCGCGCAGTGTGCTGCT----- 4009

QY 351 ypheAlaGlyValGlyLysLeuGlnGluMetAlaThrLysasnIleThraspProAlaTh 371
Db 4010 -----TCTGATGATTCGGTAGCTTCTAGCCCTGC 4038
QY 371 rLysAlaAlaValSerGlnLeuThrAsnLeuAlaGlySerAlaAlaValPheAlaGlyTr 391
||| ||| :::: |||
Db 4039 CAGTCCGCTTGTCTGACGATGCTAAAGCAAGTAGCGCCCGCATGTAGCATCCAGCGC 4098
QY 391 pThrThrAlaAlaLeuThrThrAspProAlaVal-----LysLy 404
||| ||| ||| :::: |||
Db 4099 TACAACTGCTGCCATTAGTTCGCCACATCCTTGCGCTGATCAGAGTCCACAGGGTCAAC 4158
QY 404 sAlaGluSerPheIleGlnaspThrValLysSerThrAlaSerSerThrThrGlyTyrVa 424
||| ||| ||| :::: ||| ||| |||
Db 4159 CGCTGGCTCCCATATT-----TTGCCAAGTACTGCTGGAGAGACGACAGGTAGTAT 4209
QY 424 1--AlaaspGlnThr-----ValLysLeuAlaLysThrVa 435
||| ||||| |||
Db 4210 ACCATCGGGTCAAGCGCAACACAGACGAAAGCCAAACACAGCAACACAAACGAA 4269
QY 435 LlysaspMetGlyGlyGluAlaIleThrHisThrGlyAlaSerLeuArgasnThrValas 455
||| :::: ||| ||||| ||||| |||
Db 4270 GCCAACACAAAGCCGGTCAACA-----ACCCAGACAGGT---TCATTACCGCAACGATCA 4323
QY 455 nasLeuArgGlnArgProAlaArgGluAlaaspIleGluGlyGlyThrAla 473
||| ||| :::: ||| ||| ||||| :::
Db 4324 TGCAGGAGGCAATATGCTACCGCAGACCGGCTGATGATGTGAAGCGGTACTTCT 4378

RESULT 5
US-09-287-849-21
; Sequence 21, Application US/09287849
; Patent No. US20020009459A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skelky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; FILE REFERENCE: and Their Uses
; CURRENT APPLICATION NUMBER: US/09/287, 849
; CURRENT FILING DATE: 1999-04-07
; PRIOR APPLICATION NUMBER: US 08/818, 112
; PRIOR FILING DATE: 1997-03-13
; PRIOR APPLICATION NUMBER: US 08/942, 578
; PRIOR FILING DATE: 1997-10-01
; PRIOR APPLICATION NUMBER: US 09/025, 197
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 09/056, 556
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223, 040
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 21
; LENGTH: 1801
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: tri-fusion
; OTHER INFORMATION: protein TBH9-DPV-MTI (designated Mtbf1f)
; NAME/KEY: CDS
; LOCATION: (1)..(1800)
; US-09-287-849-21
Alignment Scores:
Pred. No.: 7.24e-05 Length: 1801
Score: 143.00 Matches: 117
Percent Similarity: 34.45% Conservative: 58
Best Local Similarity: 23.03% Mismatches: 165
Query Match: 5.93% Indels: 168

DB: 10 Gaps: 25

US-09-825-414-7 (1-486) x US-09-287-849-21 (1-1801)

QY 20 ThrAlaSerAspAlaSerLeuAlaSerSerSerValArgSerValSerSerAspGln 39

Db 262 ACCGGGGGAGGAGGAGCTGACCGCGCCAGTCCGGGT----- 303

QY 40 ArgGluLeuAsnAlaIleAlaAspTyrLeuThrAspHisValPheAlaHisLysLeu 59

Db 304 -----GCTGGCGCGCGCTACGAGACGGCGTATGGCTGACG-----GTG 342

QY 60 ProProAlaAspSerAlaAspGlyGlnAla----- 69

Db 343 CCCCCCGCGGTGATCGCCGAGAACCCTGCTGAACGTGATTTCTGATAGCAGCAACCTC 402

QY 70 -----AlaValAspValHisAsnAlaGlnIleThrAlaLeuIleGlu 83

Db 403 TTGGGGCAAAACACCCCGCGGATCGCGGTACACGAGCGCGAATACGGCGAGATGGGCC 462

QY 84 ThrArgAlaSerArgLeuHisPheGluGlyGluThrProAlaThrIleAlaAspThrPhe 103

Db 463 CAAGACGCGCGCGGATGTTGGCTACGCCCGCGCGACGGCGACGGCGCGACGCTTG 522

QY 104 AlaLysAlaGluLysLeuAspArgLeuAlaThrThrThrSerGlyAlaLeu----- 120

Db 523 CTGCCGTTGAGGAGCGCGCGGAGATG-----ACCAGCGCGGCTGGCTCCTCGAGCAG 576

QY 121 -----ArgAlaThrProPheAlaMetAlaSer--LeuLeuGlnTyrMet 134

Db 577 GCCCGCGCGGTGAGAGCGCTCCGACACCGCGCGCGAACCAGTTGATGAACATGTG 636

QY 135 GlnProAlaIleAsnLysGlyAspTyrLeuProAlaProLeuLysProLeuThrProLeu 154

Db 637 CCCCAGCGCTGCAACAG-----CTGGCCACGCCACGACGCGCACACGCGCT--- 684

QY 155 IleSerGlyAlaLeuSerGlyAlaMetAspGlnValGly-----Thr 168

Db 685 ---TCTTCCAAGCTGGGTGGCTGTGGAAGACGGTCTCGCCGATCGCTCCGCATCAGC 741

QY 169 LysMetMetAspArgAlaThrGlyAspLeuHisTyrLeuSerAlaSerProAspArgLeu 188

Db 742 AACATGGTGTGATGGCCCAACACCATGTGCG--ATGACCAACTCGGGTGTGTGATG 798

QY 189 HisAspAlaMetAlaAlaSerValLysArgHisSerProSerLeuAlaArgGlnValLeu 208

Db 799 ACCAACACCTTGAGCTGATGTGAAGGGCTTGTCCGGCGCGCGCCAG----- 852

QY 209 AspThrGlyValAlaValGlnThrTyrSerAlaArgAsnAlaValArgThrValLeuAla 228

Db 853 -----GCCGTGCAAAACC--GCGCGCAAAACGGGGTCCGG----- 885

QY 229 ProAlaLeuAlaSerArgProAlaValGlnGlyAlaValAspLeuGlyValSerMetAla 248

Db 886 ---GCGATGAGCTCGGTGGGCAAGCTCGTGGGTCTTCGGGTCTGGGC----- 930

QY 249 GlyGlyLeuAlaAlaAsnAlaGlyPheGlyAsnArgLeuLeuSerValGlnSerArgAsp 268

Db 931 GGTGGGTGGCGCCCACTTGGGT-----CGGGCGGCTCGGTGGGTTCG----- 975

QY 269 HisGlnArgGlyGlyAlaLeuValLeuGlyLeuLysAspLysGluProLysAlaGlnLeu 288

Db 976 -----TTG 978

QY 289 SerGluGluAsnAspTyrPheGluAlaTyrLysAlaIleLysSerAlaSer----- 305

Db 979 TCGGTGCGCGCAGGCTGGCGCGCGCAACCAAGCAATCACCCCGCGCGCGCGCTG 1038

QY 306 -----TyrSerGlyAlaAlaLeuAsnAlaGlyLysArgMetAlaGlyLeu 320

Db 1039 CCGCTGACCAAGCTGACCAAGCGCGCGGAAAGAGGGCGCGCAGATGCTGGCGGGCTG 1098

QY 321 ProLeuAspMetAlaThrAspAlaMetGlyAlaValArgSerLeuValSerAlaSerSer 340

Db 1099 CCGGTGGG----- 1107

QY 341 LeuThrGlnAsnGlyLeuAlaLeuAlaGlyGlyPheAlaGlyVal----- 355

Db 1108 -----CAGATGGGCGCCAGGGCGCGGTGGTGGCTCAGTGTGTGCTGTTCGCCG 1161

QY 356 -----GlyLysLeu----- 358

Db 1162 CGACCTATGTGATGCCCATTTCTCCGGCAGCGCGCAAGCTTGATCCCGTGACCGGCTC 1221

QY 359 -----GlnGluMetAlaThrLysAsnIleThrAspPro--- 369

Db 1222 ATTAACACCACCTGCAATTACGGGCAAGTAGTACTGCCGTCACACGCGAGATCCGGGG 1281

QY 370 -----AlaThrLysAlaAlaValSerGlnLeuThrAsnLeuAlaGly 383

Db 1282 GCTGCCGACAGTTCAACGCTTCACCGGTGGCGCAGTCTATTGTGCGCAATTCTCCGCC 1341

QY 384 SerAlaAlaValPheAlaGlyTyrThrThrAlaAlaLeuThrThrAspProAlaValLys 403

Db 1342 GCACCGCACCTCAGCGCGCTGCCATGCGCGCAATTG-----CAAGCTGTGCCG 1392

QY 404 LysAlaGluSerPheIleGlnAspThrValLysSerThrAlaSerSerThrThrGlyTyr 423

Db 1393 GGGCGGACACAGTACATC--GGCCTGTGACGTGCGTGGCGCTCTGCCAACAATAT 1449

QY 424 ValAlaAspGlnThrValLysLeuAlaLysThrValLysAspMetGlyGlyAlaIle 443

Db 1450 GAG-----CTCATGACGATTAATTACAGTTCGGGAGCGTCGAC 1488

QY 444 ThrHisThrGlyAlaSerLeuArg 451

Db 1489 GCTCAT--GGCGCATGATCCGC 1509

RESULT 6

US-09-287-849-1

; Sequence 1, Application US/09287849

; Patent No. US20020009459A1

; GENERAL INFORMATION:

; APPLICANT: Reed, Steven G.

; APPLICANT: Skeiky, Yasir A.W.

; APPLICANT: Dillon, Davin C.

; APPLICANT: Alderson, Mark

; APPLICANT: Campos-Neto, Antonio

; APPLICANT: Corixa Corporation

; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens

; FILE REFERENCE: 014058-009020US

; CURRENT APPLICATION NUMBER: US/09/287, 849

; CURRENT FILING DATE: 1999-04-07

; PRIOR APPLICATION NUMBER: US 08/818,112

; PRIOR FILING DATE: 1997-03-13

; PRIOR APPLICATION NUMBER: US 08/942,578

; PRIOR FILING DATE: 1997-10-01

; PRIOR APPLICATION NUMBER: US 09/025,197

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 09/056,556

; PRIOR FILING DATE: 1998-04-07

; PRIOR APPLICATION NUMBER: US 09/223,040

; NUMBER OF SEQ ID NOS: 46

; SOFTWARE: Patentln Ver. 2.1

; SEQ ID NO 1

; LENGTH: 2287

; TYPE: DNA

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence:tr1-fusion

; OTHER INFORMATION: protein Ral2-TbH9-Ra35 (designated Mb32-Mtb39

; NAME/KEY: modified_base

; LOCATION: (30)

;; PRIOR APPLICATION NUMBER: US/09/177,349
;; PRIOR FILING DATE: 1998-10-23
;; NUMBER OF SEQ ID NOS: 14
;; SOFTWARE: PatentIn Ver. 2.0
;; SEQ ID NO 2
;; LENGTH: 5036
;; TYPE: DNA
;; ORGANISM: Mycobacterium tuberculosis
US-09-918-951-2

Alignment Scores:
Pred. No.: 0.00239 Length: 5036
Score: 133.50 Matches: 108
Percent Similarity: 33.41% Conservative: 43
Best Local Similarity: 23.89% Mismatches: 182
Query Match: 5.53% Indels: 120
DB: 10 Gaps: 16

US-09-825-414-7 (1-486) x US-09-918-951-2 (1-5036)

QY 64 SerAlaAspGlyGlnAlaAlaValAspValHisAsnAla----- 76
Db 97 AGCGTCGATGGCAGACGCCCGGTCACCGCGCGCAACAGCTTCTTCGATTGAAGGGAATG 156
QY 77 GlnIleThrAlaLeuIleGluThrArgAlaSerArgLeuHisPheGluGlyGluThrPro 96
Db 157 AAGATGACCTCGCTTATCGATTACATCTGAGCTGTTCGCGACGAGACGCCCGCCCG 216
QY 97 AlaThrIleAlaAspThrPheAlaLysAlaGluLysLeuAspArgLeuAlaThrThr 116
Db 217 TCGTTCGTGGC-----GCTCCGGGACGGGCCATGACCACT 252
QY 117 SerGlyAlaLeuArgAlaThrProPheAlaMetAlaSerLeuLeuGlnTyrMetGlnPro 136
Db 253 GCCGGGCTGATCGATATCGCGCGCCGACCAATCTCATCGGTGGCGGCCCAATGTGGTGGC 312
QY 137 AlaIleAsnLysGlyAspTrpLeuProAlaProLeuLysProLeuThrProLeuIleSer 156
Db 313 GGTCTGAATCTGGGTGCCGGGACCC----- 339
QY 157 GlyAlaLeuSerGlyAlaMetAspGlnValGlyThrLys-MetMetAspArgAlaThrGln 176
Db 340 -----ATGAGCGGATTCGGGACGCGCGCTGCGCGCTGGCATGGCTTTGGCGACGAGCTC 393
QY 176 YAspLeuHisTyrLeuSerAlaSerProAspArgLeuHisAspAlaMetAlaAlaSerVal 196
Db 394 GCCA-----ATGTCGGCTTCGCGCGGTGACGCGGCGGGGGTGGCAAGCTCATC 444
QY 196 LlysArgHisSerProSer-----IeuAlaArgGlnValLeuAspThrGlyValAlaVala 214
Db 445 ACGACCGATGTGGTGGCGGCTGCGGTAGCGGACTGGGTGCTGGGTCTGGGTCAAGGT 504
QY 214 GlnThrTyrSerAlaArgAsnAlaValArgThr----- 225
Db 505 GGCCTGGCTTCGCCCGCGCTCAAGCGGTGTTTCGGCGGTCAAGTCCGCTTGGCTGCCAG 564
QY 226 -----ValLeuAlaProAlaLeuAlaSerArgProAlaValGlnGlyAlaVala 242
Db 565 GTCGGTCTGGGTTTACTGCCGTGATTG-----AGGCCGAG-GTCGGCGCTCAGGTTGG 617
QY 242 pLeuGlyValSerMetAlaGlyGlyLeuAlaAlaAsnAlaGlyPheGlyAsnArgLeuLe 262
Db 618 TGGTGGTTAGGTATTGGACGGGTCTGGGTCTCAGCGCGGTATGGC----- 666
QY 262 uSerValGlnSerArgAspHisGlnArgGlyGlyAlaLeuValLeuGlyLeuLysAspLys 282
Db 667 -----TTTGGCGCGCGGGGTTGGCTGGCTGGCTGGGTGT-- 699
QY 282 sGluProLysAlaGlnLeuSerGluGluAsnAspTrpLeuGluAlaTyrLysAlaIleLys 302
Db 700 -----CAGCGCGCGGTGTGATCGG 719
QY 302 sSerAlaSerTyrSerGlyAlaAlaLeuAsnAlaGlyLysArgMetAlaGlyLeuProLe 322

Db 720 TGGAGCGCGCGCGGCTATCGGTGCCGCGCTCGCGCGCTTAGCGGCAATGCC-- 777
QY 322 uAspMetAlaThrAspAlaMetGlyAlaValArgSerLeuVal----- 336
Db 778 -CAGATCGAGTTGCCGCCAGGCGCTGGCTGGCTGGTTCGGCGCTGTGTGGCGCG 836
QY 337 -SerAlaSerSerLeuThrGlnAsnGlyLeuAlaLeuAlaGlyPheAlaGlyValGln 356
Db 837 CCAGCGCGGCATCGCTAGCCAGATCGGTCTCTCAGCCGGGTGGTGGCTCGCGCGCTCGG 896
QY 356 YLys----- 357
Db 897 CAATGTACCGCGCTGACCGCGGTCAGCAACGACGAGTGTGGCTTCCACGCAAGCGG 956
QY 358 -----LeuGlnGluMetAlaThrLysAsnIle 366
Db 957 CCAGCGCGGCTGATCGCCAGTGAAGCGCTGCCCTTGAACGGCGCTGCTATGCCCTCATCT 1016
QY 366 eThrAspProAlaThrLysAlaAlaValSerGlnLeuThrAsnLeuAlaGlySerAlaAl 386
Db 1017 GTCGGCGCGCTTAGCCGGTGTGCGGTGGGTGTGTGTCAGCGCGCGCGCGCTGGCGCGCG 1076
QY 386 aValPheAlaGlyTyrThrThrAlaAlaLeuThrThrAspProAlaValLysLysAlaGln 406
Db 1077 GTTGGGCTTGGGA---GCGGTGCGGACCCGACTCTCAGCGCGCGCGCTGGCGCGCG 1133
QY 406 uSerPheIleGlnAspThrValLysSerThrAlaSerSerThrThrGlyTyrValAlaAs 426
Db 1134 TGGCGTGTGGCCAAAGACC-----GAGCGGCGCTGCTGAGTGTGGCGGGGTCCGCGG 1187
QY 426 pGlnThrValLysLeuAlaLysThrValLysAspMetGlyGlyGluAlaIleThrHisTh 446
Db 1188 GGCAACC-----GCGGCGGGGTGCGGGGACACAGCGGACATCTGGGGCCACGA 1238
QY 446 rGlyAlaSerLeuArg-----AsnThrValAsn 455
Db 1239 GGGAGCCGCACTGGGCACTGTGTCAGACAGGTCAAC 1272

RESULT 8
US-09-815-242-4584
; Sequence 4584, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011a
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 4584
; LENGTH: 6258
; TYPE: DNA
; ORGANISM: Staphylococcus aureus
US-09-815-242-4584

Alignment Scores:
Pred. No.: 0.00501 Length: 6258
Score: 131.50 Matches: 125
Percent Similarity: 35.34% Conservative: 75
Best Local Similarity: 22.08% Mismatches: 227
Query Match: 5.45% Indels: 139
DB: 10 Gaps: 27

US-09-825-414-7 (1-486) x US-09-815-242-4584 (1-6258)

QY 3 IleasnArgargValGln---GlnProProValThrAlaThrAspSerPheargThrAla 21
||||| ||||| ||||| |||||
Db 553 ATTACAATGCAGTTCAAGTTGCTAATAAACGTACTGCACAGATTAATAATGCGACAGCA 612
QY 22 SerAspAlaSerIleAlaSerSerSerValArgSerVal----- 34
||| ||| |||
Db 613 ATGCCTACTAATTAGCTGGTGTGACACAAACGACATTCCTGTGACAGTAATCAAT 672
QY 35 -----SerSerAspGlnGlnArgGluIle 42
||| ||| |||
Db 673 GATGGCAGTACTGAAGAAGTACACAGAGCTATTTTCACAAAAGCGGATAAACGTGAGTTA 732
QY 43 AsnAlaIleAlaSerPyrIleuThrAspHisValPheAlaAlaHisLysLeuProAla 62
||||| ||||| ||||| |||||
Db 733 ATCACAGCTAAAAATCATTTAGATGATCCAGTACAGTACAGATGGTAAAAAGCCAGGTACA 792
QY 63 AspSerAlaAspGlyGlnAlaAlaValaLysPvalHisAsnAlaGln-----IleThr 79
||| ||| |||
Db 793 ATTACGCGTACCAATATGCA-----ATTCATTAATGCGCAACACAAATCAATACT 843
QY 80 AlaLeuIleGluThrArgAlaSerArgLeuHisPheGluGlyGluThrProAlaThrIle 99
||| ||| |||
Db 844 GCGAAAAACAGAGCGCAA---CAAGTTATTAAATGAGCGGTGCAACACCAACAAGTT 900
QY 100 AlaaspThrPheAlaLysAlaGluLysLeuAspArgLeuAlaThrThrThrSerGlyAla 119
||| ||| |||
Db 901 TCTGACGCTTTAATTAAGTT-----CGTGCTGCACAAACTAAGATAAATGAA 948
QY 120 LeuArgAlaThrProPheAlaMetAlaSerLeuGln----- 132
||| ||| |||
Db 949 GCTAAAGCA-----TTACTTCAAAATAAAGAGATAATAGCCAA 987
QY 133 -----TyrMetGlnProAlaIleAsnLysGlyAspTyrLeuProAla 146
||| ||| |||
Db 988 TTAGTACATCTTAAAAATAACTTACAAAAGTTCTGTGAACCAA----- 1029
QY 147 ProLeuLysProLeuThrProLeuIleSerGlyAlaLeuSerGlyAlaMetaspGlnVal 166
||| ||| |||
Db 1030 -----GTACCATCAACTACTGTGTATGACGCAACAAGTATGTATACTAT 1074
QY 167 GlyThrLysMetMetaspArgAlaThrGlyAspLeuHisTyrLeuSerAlaSerProasp 186
||| ||| |||
Db 1075 AATGCGAAGAAACGCGAAGCAGAACT-----GAAATAACTGCAGCTCAACGT 1122
QY 187 ArgLeuHis-----AspAlaMetAlaAlaSerValLysArgHisSerProSerLeuAla 204
||| ||| |||
Db 1123 GTTATGTGACAATGTTGACTGCACTGCACACAACAATTTCAGATGAATAACACTACAGCACA 1182
QY 205 ArgGlnValLeuAspThrGlyValAlaValaGlnThrTyrSerAlaArgAsnAlaValArg 224
||| ||| |||
Db 1183 CGCAATGATTTAACAAATCAAAATTTCACAAGCTACAAACTTAGCT-----GCTGTGAA 1236
QY 225 ThrValLeuAlaProAlaLeuAlaSerArgProAlaValGlnGlyAlaValaLysPleuGly 244
||| ||| |||
Db 1237 TCTGTTAAGCAAAAGTGG---AATAGTTTAGATGCTGCAATGGTAACTTACAAACGGCT 1293
QY 245 ValSerMetAlaGlyGlyLeuAlaAlaAsnAlaGlyPheGlyAsnArgLeuLeuSerVal 264

Db 1294 ATCAACGATAAATCTGGAACATTTAGCGAGCCAAAACCTTC----- 1332
QY 265 GlnSerArgaspHisGlnArgGlyGlyAlaLeuValLeuGlyLeuLysaspLysGluPro 284
||| ||||| |||
Db 1333 CTAGATGCTGATGAGCAAAAACGTAATGCTTACAAATCAAGCTGTATGCAAAATGCCGAA--- 1389
QY 285 LysAlaGluLeuSerGluGluAsnAspTyrPleuGluAlaTyrLysAlaIleLysSerAla 304
||| ||| ||| |||
Db 1390 ---ACAATTTTAATAAACAACACTGACCGCAATACAGCGAAAACCTGCAGTCAACAAGCA 1446
QY 305 SerTyrSerGlyAlaAlaLeuAsnAlaGlyLysArgMetAlaGlyLeuProLeuAspMet 324
||| ||| |||
Db 1447 CTTAATAAT-----GTTAATAGTCCGAACAACATGCATTAATGTGTACGCAAAACTTA 1497
QY 325 AlaThrAspAlaMetGlyAlaValaLysSerLeuValSerAlaSerSerLeuThrGlnAsn 344
||| ||| ||| |||
Db 1498 AATAATGCGAACAAGCAGCGATTTACAGCAATCAATGGCGCATCTGATTTAAATCAAAAA 1557
QY 345 Gly-----LeuAlaLeuAlaGlyGlyPheAlaGlyValGlyLysLeuGlnGlu 360
||| ||| |||
Db 1558 CAAAAGATGCATTAAGACACACAAGCTAATGGTCTCAACGCGTATCTTAATGCACAAGAT 1617
QY 361 MetAlaThrLysAsnIleThrAspProAlaThr-----LysAlaAlaVal 375
||| ||| ||| |||
Db 1618 GTA---CAACGTAAATGCGAAGTGAACGTGAACGCGCAATGGGCACATTAACAATGCCATC 1674
QY 376 SerGlnLeuThrAsn---LeuAlaGlySerAlaAlaValPheAlaGlyTyrThrThrAla 394
||| ||| ||| |||
Db 1675 GCAGATAAGACGAATACGTTAGCAAGCAGTAATATGTTAACGCCGATAGCACTAAACAA 1734
QY 395 ---AlaLeuThrThrAspProAlaValLysLysAlaGluSerPheIleGln-----Asp 411
||| ||| ||| |||
Db 1735 AATGCTTACACAACT-----AAAGTTACCAATGCTGAACATATATATAGCGGTACGCCA 1788
QY 412 ThrValLysSerThrAlaSerSerThrThrGlyTyrVal----- 424
||| ||| ||| |||
Db 1789 ACGGTGTATACACACCCTTCAGAAGTACAGCTGCAGCTAATCAAGTAACAACAGCGGAAA 1848
QY 425 -----AlaAspGlnThrValLysLeuAlaLys----- 433
||| ||| ||| |||
Db 1849 CAAGAAATTAATGGTGACGAAAGATTACGTGTGCCAAAAACAACGCCAATACTGCTATT 1908
QY 434 -----ThrValLysaspMetGlyGlyGlu 441
||| ||| ||| |||
Db 1909 GATGCATTAAACGAATTAAATACACCTCAAAAAGCTAAATTAAAGAACAAGTGGGACAA 1968
QY 442 Ala-----IleThrHisThrGlyAlaSerLeuArgAsnThr 453
||| ||| ||| |||
Db 1969 GCCAATAGATTAGAGACATACAAACTGTTCAACAACAATGGACAAGCATGACAACATGCA 2028
QY 454 ValAsnAsnLeuArgGlnArgProAlaArgGluAlaAlaAspIleGlu-----GluGlyGly 471
||| ||| ||| |||
Db 2029 ATGAAGAGCTTAAGAGATAGTATGTCTAACGAACAACAACAGTCAAGCAAGTCAAAACTAT 2088
QY 472 ThrAlaAlaSerProSer 477
||| ||| ||| |||
Db 2089 ACAGACGCAAGTCCGAAT 2106

RESULT 9
US-09-815-242-8512
; Sequence 8512, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes In

```

: TITLE OF INVENTION: Prokaryotes
: FILE REFERENCE: ELITRA.011A
: CURRENT APPLICATION NUMBER: US/09/815,242
: CURRENT FILING DATE: 2001-03-21
: PRIOR APPLICATION NUMBER: 60/191,078
: PRIOR FILING DATE: 2000-03-21
: PRIOR APPLICATION NUMBER: 60/206,848
: PRIOR FILING DATE: 2000-05-23
: PRIOR APPLICATION NUMBER: 60/207,727
: PRIOR FILING DATE: 2000-05-26
: PRIOR APPLICATION NUMBER: 60/242,578
: PRIOR FILING DATE: 2000-10-23
: PRIOR APPLICATION NUMBER: 60/253,625
: PRIOR FILING DATE: 2000-11-27
: PRIOR APPLICATION NUMBER: 60/257,931
: PRIOR FILING DATE: 2000-12-22
: PRIOR APPLICATION NUMBER: 60/269,308
: PRIOR FILING DATE: 2001-02-16
: NUMBER OF SEQ ID NOS: 14110
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 8512
: LENGTH: 17388
: TYPE: DNA
: ORGANISM: Staphylococcus aureus
: FEATURE:
: NAME/KEY: CDS
: LOCATION: (1)...(17388)
: US-09-815-242-8512

```

Alignment Scores:

Pred. No.:	0.0197	Length:	17388
Score:	131.50	Matches:	125
Percent Similarity:	35.34%	Conservative:	75
Best Local Similarity:	22.08%	Mismatches:	227
Query Match:	5.45%	Indels:	139
DB:	10	Gaps:	27

US-09-825-414-7 (1-486) x US-09-815-242-8512 (1-17388)

```

QY 3 1leasnargargvalgn--GlnProProvalThralathraspserpheargThrala 21
    |||||  |||||  |||||  |||||
Db 7384 ATTAACAATGCAGTTCAGTTGCTAATAAACGTACTGCACAGATTAATAATGCACAGCA 7443
QY 22 SeraspalaserleualaserSerSerValArgSerVal----- 34
    ::|||  ::|||  ::|||
Db 7444 ATGCCCTACTAATTAGCTGCTGCTAGCACACAGCAGATTCCCTGTGACAGTAACCTACAAT 7503
QY 35 -----SerSeraSpGlnGlnArgGluIle 42
    ::|||  ::|||  ::|||
Db 7504 GATGGCAGTACTGAGAAGTACAGAGCTATTTCACAAAAGCGCATTAACGTGAGTTA 7563
QY 43 AsnAlaIleAlaAspTyrLeuThrAspHisValPheAlaAlaHisLysLeuProAla 62
    ::|||  ::|||  ::|||  |||
Db 7564 ATCACAGCTAAATAATCATTTAGATGATCCAGTGAGTACAGATGTTAAAGCCAGGTACA 7623
QY 63 AspSeraIaaspGlyGlnAlaAlaValAspValHisAsnAlaGln-----IleThr 79
    ::|||  ::|||  ::|||  |||
Db 7624 ATTACGCACTACAATAATGCA-----ATTCTAATGCGCAACAACAATCAATACT 7674
QY 80 AlaIleuIleGluThrArgAlaSerArgLeuHisPheGluGlyIleThrProAlaThrIle 99
    |||  |||  ::|||  |||  |||  |||
Db 7675 GCGAAAACAGAAAGCGCAA---CAAGTTATTATAATGAGCGTGCACACACACAACAAGTT 7731
QY 100 AlaAspThrPheAlaIleAlaGluLysLeuAspArgLeuAlaThrThrThrSerGlyAla 119
    ::|||  |||  |||  |||  |||  |||
Db 7732 TCTGACGCTTTAATAAGTT-----CGTGCTGCACAACATAAGATAATGAA 7779
QY 120 LeuArgAlaThrProPheAlaMetAlaSerLeuLeuGln----- 132
    ::|||  |||  |||  |||  |||
Db 7780 GCTAAAGCA-----TTACTTCAAAATAAAGAAAGATAATAGCCAA 7818
QY 133 -----TyrMetGlnProAlaIleAsnLysGlyAspTrpLeuProAla 146
    ::|||  ::|||  ::|||  ::|||

```

```

Db 7819 TTAGTAACATCTAAATAACTTACAAAGTCTGTGACCAA----- 7860
QY 147 ProleuLysProleuThrProleuIleSerGlyAlaLeuSerGlyAlaMetAspGlnVal 166
    |||  ::|||  ::|||
Db 7861 -----GTACCATCACTACTGTTATGACGCAACAAGTATTGATTAACATA 7905
QY 167 GlyThrLysMetMetAspArgAlaThrGlyAspLeuHisTyrLeuSerAlaSerProAsp 186
    |||  ::|||  |||  ::|||
Db 7906 AATGCGAAGAAACGCGAAGCAGAAACT-----GAATTAAGTCAGCTCAACGT 7953
QY 187 ArgLeuHis-----AspAlaMetAlaAlaSerValLysArgHisSerProSerLeuAla 204
    ::|||  |||  |||  ::|||
Db 7954 GTTATTGACAATGTTGATGCACACTGCACACAATAATTTCAGATGAACACTACAGCACAA 8013
QY 205 ArgGlnValLeuAspThrGlyValAlaValGlnThrTyrSerAlaArgAsnAlaValArg 224
    |||  |||  ::|||  |||  |||  |||
Db 8014 CCGAATGATTTTAACAATCAATTTTCACAAGCTACAACTTAGCT-----GCTGTTGAA 8067
QY 225 ThrValLeuAlaProAlaLeuAlaSerArgProAlaValGlnGlyAlaValAspLeuGly 244
    ::|||  |||  |||  |||  |||
Db 8068 TCTGTTAAGCAAAAGTGCG--AATAGTTAGATGTTGCAATGGGTAACCTTACAAACGGCT 8124
QY 245 ValSerMetAlaGlyGlyLeuAlaAlaAsnAlaGlyPheGlyAsnArgLeuLeuSerVal 264
    ::|||  |||  |||  |||  |||
Db 8125 ATCAACGATAAATCTGGAACATTTAGCGAGCCAAAACCTTC----- 8163
QY 265 GlnSerArgAspHisGlnArgGlyGlyAlaLeuValLeuGlyLeuLysAspLysGluPro 284
    |||  |||  |||  |||  |||
Db 8164 CTAGATGCTGATGAGCAAAAACGTAATGCTTACAATCAAGCTGTATCAAAATGCCGAA--- 8220
QY 285 LysAlaGlnLeuSerGluGlnAsnAspTrpLeuGluAlaTyrLysAlaIleLysSeraIa 304
    |||  ::|||  ::|||  |||  |||  |||
Db 8221 ---ACAATTTTAATAAACAACACTGCACCGAATACAGCGAAAACTGCAGTCGAACAGCA 8277
QY 305 SerTyrSerGlyAlaAlaLeuAsnAlaGlyLysArgMetAlaGlyLeuProleuAspMet 324
    ::|||  ::|||  ::|||
Db 8278 CTTAATAAT-----GTTAATAAGTCGGAACATGCATTAAATGTACGCAAAACTTA 8328
QY 325 AlaThrAspAlaMetGlyAlaValArgSerLeuValSerAlaSerSerLeuThrGlnAsn 344
    |||  ::|||  ::|||  |||  |||  |||
Db 8329 AATAATGCCGAACAACAGCAGCATTTACAGCAATCAATGCCCATCTGATTTAAATCAAAA 8388
QY 345 Gly-----LeuAlaLeuAlaGlyGlyPheAlaGlyValGlyLysLeuGlnGlu 360
    |||  |||  |||  |||  |||
Db 8389 CAAAAAGATGCATTAAAAGCACAGCTAATGTTGCTCAACGCCGTATCTAATGCACAGAT 8448
QY 361 MetAlaThrLysAsnIleThrAspProAlaThr-----LysAlaAlaVal 375
    ::|||  ::|||  |||  |||
Db 8449 GTA---CAACGTAATGCGACTGCACTGAACACGCGCAATGGGCACATTAAACATGCCATC 8505
QY 376 SerGlnLeuThrAsn---LeuAlaGlySeraIaAlaValPheAlaGlyTrpThrThrAla 394
    ::|||  |||  |||  |||  |||  |||
Db 8506 GCAGATAAGACGAATAGCTTAGCAGACAGTAATATGTTAAAGCCGATAGCACTTAACAA 8565
QY 395 ---AlaLeuThrThrAspProAlaValLysLysAlaGluSerPheIleGln-----Asp 411
    |||  |||  |||  |||  |||
Db 8566 AATGCTTACACAACCT-----AAAGTTACCAATGCTGAACATATTATTAGCGGTACGCCA 8619
QY 412 ThrValLysSerThrAlaSerSerThrThrGlyTyrVal----- 424
    |||  ::|||  |||  |||
Db 8620 ACCGTTGTTACACACCTTCAGAGAGTAACAGCTGCAGCTTAATCAAGTAACAGCGCGAAA 8679
QY 425 -----AlaaspGlnThrValLysLeuAlaLys----- 433
    |||  ::|||  ::|||  |||
Db 8680 CAAGAAATTAAATGTTGACGAGAAAGATTACGTTGCAAAACAAACGCCAATACTGCTATT 8739
QY 434 -----ThrValLysAspMetGlyGlyGlu 441
    |||  ::|||  ::|||
Db 8740 GATGCATTAACGCAATTAAATACACCTCAAAAAGCTAAATTAAGAACAGCAAGTGGGACAA 8799
QY 442 Ala-----IleThrHisThrGlyAlaSerLeuArgAsnThr 453
    |||  ::|||  ::|||  |||
Db 8800 GCCAATAGATTAGAAGACATACAAACTGTTCAAACAAATGGACAGCATTTGAACAATGCA 8859

```


QY 454 ValAsnAsnLeuArgGlnArgProAlaArgGluAlaAspIleGlu-----GluGlyGly 471
::: ||||| ||| ||| ::::: :::
Db 8860 ATGAAGGCTTAGAGATAGTATTGCTTAACGAACAACAGTCAAAAGTCAAAACTAT 8919
QY 472 ThrAlaAlaSerProSer 477
||| ||||| ||| :::
Db 8920 ACAGACGCAAGTCCGAAT 8937
RESULT 10
US-09-815-242-8615
; Sequence 8615, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8615
; LENGTH: 7035
; TYPE: DNA
; ORGANISM: Staphylococcus aureus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(7035)
US-09-815-242-8615
Alignment Scores:
Pred. No.: 0.00915 Length: 7035
Score: 129.50 Matches: 86
Percent Similarity: 39.55% Conservative: 107
Best Local Similarity: 17.62% Mismatches: 204
Query Match: 5.37% Indels: 91
DB: 10 Gaps: 18
US-09-825-414-7 (1-486) x US-09-815-242-8615 (1-7035)
QY 8 GlnGlnProProValThrAlaThrAspSerPheArgThrAlaSerAspAlaSerIleuAla 27
||| ||||| ::||| ::||| ||| :::::
Db 2290 CAACAATCACAAGGTGTGTCACAACAAGTAAAGCTGACTGCACAAAGTGTCTTACGAGCACA 2349
QY 28 SerSerSerValArgSerValSerSerAspGlnGlnArgGluIleAsnAlaIleAlaasp 47
||| ||||| ::||| ::||| ||| :::::
Db 2350 TCTGGTTCAATATGACTTCACATCAGCTAGTACTCGAAATCGACAAGTGTAAAGC--- 2406
QY 48 TyrLeuThrAspHisValPheAlaAlaHisLysLeuProProAlaaspSerAlaaspGly 67
||| ||||| ||| ||||| ||| :::::
Db 2407 ---CTATCTGATTCTGTGAGTGCATCTAAGTCAATTAAGCACACATCTGAAGTAAAT----- 2457

QY 68 GlnAlaAlaValAspValHisAsnAlaGlnIleThrAlaLeuIleGluThrArgAlaSer 87
||| ::::: ||| ||||| ||| :::::
Db 2458 -----AGTGTATCAAGCTCAACAAGCAGACAGTTTAACTGAATTCACAAGTGTGA 2505
QY 88 ArgLeuHisPheGluGlyGlu---ThrProAlaThrIleAlaaspThrPhe----- 103
||| ||| ::||| |||
Db 2506 TCATCAAGCATGTGCAGGTTCAGTTAGTAAATCAACATCATTAAGCGATTTTATTCGAAC 2565
QY 104 --AlaLysAlaGluLysLeuAspArgLeuAlaThrThrThrSerGlyAlaLeuArgAla 122
::: ||||| ::||| ::::: ||||| ||| ::||| |||
Db 2566 TCTAGCAGTACTGAAAAATCAGAGAGTGTCTCAACAAGTACATCTGATTCATTCGCTACA 2625
QY 123 ThrProPheAlaMetAlaSerLeuLeuGlnTyrMetGlnProAlaIleAsnLysGlyAsp 142
::: |||
Db 2626 TCA----- 2628
QY 143 TripleuProAlaProLeuLysProLeuThrProLeuIleSerGlyAlaLeuSerGlyAla 162
||| ||||| ::||| ::||| ||||| |||
Db 2629 -----ACATCGTTAAGTATTCAGTACGATGAGTACATGAGTACATCAGAGAAC 2670
QY 163 MetAspGlnValGlyThrLysMetMetAspArgAlaThrGlyAspLeuHisTyrLeuSer 182
::: ||| ::||| |||
Db 2671 TTGTCTTAAG-----TCACAAGCTTATCAACAAGTACATCCGAT-----TCG 2712
QY 183 AlaSerProAspArg--LeuHisAspAlaMetAlaAlaSerValLysArgHisSerPro 201
||| ||||| ::||| ||||| |||
Db 2713 GCTAGTACATCACAATCGGTAGTGTAGTATGATCAATCAATTCATTT----- 2757
QY 202 SerLeuAlaArgGlnValLeuAspThrGlyValAlaValGlnThrTyrSerAlaArgAsn 221
||| ::||| ::||| ||||| |||
Db 2758 AGTACATCAGATCATTAAGTGTAGTCAAGTACATGATCAATCTAATAGTATTTCTTAAT 2817
QY 222 AlaValArgThrValLeuAlaProAlaLeuAlaSerArgProAlaValGlnGlyAlaVal 241
::||| ::||| ::||| |||
Db 2818 AGCATTTTCAATTCAGTGTAGTGCATCAACAAGCAATTTGGAATCACAAGTACATCATATA 2877
QY 242 AspLeuGlyValSerMetAlaGlyLysLeuAlaAlaAsnAlaGlyPheGlyAsnArgLeu 261
||| ||| ::||| |||
Db 2878 TCATTTAAGTACAAGTGATTTCTAAATCGATGAGTACATCAGAAATCATTTAAGCGAT----- 2931
QY 262 LeuSerValGlnSerArgAspHisGlnArgGlyGlyAlaLeuValLeuGlyLeuLysAsp 281
||| ||| ::||| |||
Db 2932 ---TCAACGAGCACAAAGTGAATTCAGTATCTGATCATTAAGTGTAGCGGG----- 2979
QY 282 LysGluProLysAlaGlnLeuSerGluGluAsnAspTyrLeuGluAlaTyrLysAlaIle 301
::||| ::||| |||
Db 2980 -----TCACAAGGTGTCTCAACAAGCACACATCAGACGTCATGAATGATGA 3033
QY 302 Lys---SerAlaSerTyrSerGlyAlaAlaLeuAsnAlaGlyLysArgMetAlaGlyLeu 320
||| ||| ||||| ||| ::||| |||
Db 3034 AGCGATTCTATGAGTACAAAGTGTATCATTTAGCTGCATCAGACAGTAATCATATGCCGTA 3093
QY 321 ProLeuAspMetAlaThrAspAlaMetGly-----AlaValArgSerLeuVal 336
||| ||||| |||
Db 3094 AGTAGTTTCGATGAGCACGCTTCAGTCAGGTAGTACATCAGAAATCATTTAAGTATTCGATA 3153
QY 337 SerAlaSerSerLeuThrGlnAsnGlyLeuAlaLeuAlaGlyGlyPheAlaGlyValGly 356
||| ||| ||| ||||| |||
Db 3154 AGTACATCGGATTCTGACAGTAAATCATTTATCATTTGAGTACAAAGTCAAGTCTGGC----- 3207
QY 357 LysLeuGlnGluMetAlaThrLysAsnIleThrAspProAlaThrLysAlaAlaValSer 376
||| ::||| |||
Db 3208 -----TCACAAGCACATCAACATCAACAAGTAGTTTCAAGTACGTAATGTCG 3252
QY 377 GlnLeuThrAsnLeuAlaGlySerAlaAlaValPheAlaGlyTyrThrThrAlaAlaLeu 396
::: ||| ::||| ||||| |||
Db 3253 GAATCACAAGTACATCTGTGTTCAATGAGCACACAAGTCAATCCGATTCAACAAGCATATCA 3312
QY 397 ThrThrAspProAlaValLysLysAlaGluSerPheIleGlnAspThrValLysSerThr 416
||| ||| ||||| |||
Db 3313 ACG-----TCGTTAGTGAATTCACAGGAGTGAATGACAAA 3345

QY 417 AlaserSerThrThrGlyTyrValAlaaspGlnThrValIlysLeuAlaIysThrValIlys 436
Db 3346 TCAGCTTCAACTGCA-----TCAGTGAATCAATATCACAAAGTGATCA 3390
QY 437 ---AspMetGlyGlyGluAlaIleThrHisThrGlyAlaSerLeuArgAsn----- 452
Db 3391 ACAAGTACATCTGTTCCGTTAAGTACTTCGACATCGTTAAGTACAAGTAATTCAGAACGC 3450
QY 453 -----ThrValAsnAsnLeuArgGlnArgProAlaArgGluAlaAspIleGlu 469
Db 3451 ACATCAACATCTATGAGTGAATTCACACAGCTTAAGTACATCAGAGTCTGATTCACGAGC 3510
QY 470 GlyGlyThrAlaAlaAlaSerProSer 477
Db 3511 GACTCAACGTCGACGAGCGATTCT 3534

RESULT 11
US-09-815-242-4780
; Sequence 4780, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4780
; LENGTH: 7302
; TYPE: DNA
; ORGANISM: Staphylococcus aureus
US-09-815-242-4780

Alignment Scores:
Pred. No.: 0.0108 Length: 7302
Score: 129.00 Matches: 112
Percent Similarity: 35.39% Conservative: 89
Best Local Similarity: 19.72% Mismatches: 237
Query Match: 5.35% Indels: 130
DB: 10 Gaps: 23

US-09-825-414-7 (1-486) x US-09-815-242-4780 (1-7302)
QY 3 IleasnarGArgValGlnGlnProValThrAlaThrAspSerPheArgThrAlaSer 22
Db 274 TTTAAACCAAAACAAAGATGCATTAAAAACACACAAGTTAACATGCACAAACGTGTATCT 333
QY 23 AspaIaSerLeuAlaSerSerValArgSerValSerSerAspGlnArgGluIle 42
Db 334 GATGCAATAATACGTTCAACACACTGCAACTGAATTGACACAGTGGATGACAGCAGCTTTAA 393

QY 43 AsnAlaIleAlaAspTyrLeuThrAspHisValPheAlaAlaHisIlysLeuProProAla 62
Db 394 GCAGCTATTGCTGAT-----AAAGAAAGAACAAAGCAAGCGGTAATTATGTCAATGCT 447
QY 63 AspSerAlaAspGlyGlnAla-----AlaVal 71
Db 448 GATCAAGAAAAACGTCAAGCGTATGATTCAAAAGTGACTTAACGCTGAATAATATCATTAGT 507
QY 72 AspValHisAsnAlaGlnIleThr-----AlaLeuIleGluThrArg 85
Db 508 GGTACACCGAATGCGACATTTAACAGTCAATGACGTAATAGTGGCGCATCACAAGTCAAT 567
QY 86 AlaserArgLeuHisPheGluGlyGluThrProAlaThrIleAlaAspThrPheAlaIys 105
Db 568 GCGGCTAAACACAGCATTAATGGTGATTAACAACACTTACGTGTAGCGAAAGACATGCCAAC 627
QY 106 AlaGluLysLeuAspArgLeuAlaThrThrThrSerGlyAlaLeuArgAlaThrProPhe 125
Db 628 -----AATCAATTTGACGGCTTAGCACAAATTGATAATGCA 663
QY 126 AlaMetAlaSerLeuGlnIleTyrMetGlnProAlaIleAsnLysGlyAspTrpLeuPro 145
Db 664 CAAAAAGCAAAATTTAAAGAACACAAAGTTCAAAGTGCACATCATTTAGATGGT-----GTT 717
QY 146 AlaProLeuLysProLeuThrProLeuIleSerGlyAlaLeuSerGlyAlaMetAspGln 165
Db 718 CAACTGTTAAATAATAGTTCTCAAAACGTTGAATACAGCGATGAAGGCTTAAAGATAGT 777
QY 166 ValGlyThrLysMetMetAspArgAlaThrGlyAspLeuHisTyrLeuSerAlaSerPro 185
Db 778 ATTGCAATGAAGCAACAATTAAAGCA-----GGTCAAAACTACACTGACGCAAGTCCA 831
QY 186 AspArgLeuHisAsp-----AlaMetAlaAlaSerValLysArgHis 199
Db 832 AATAATCGTAACGAGTACGACAGTGCAGTTACTGCGCAAAAGCAAAATCATTAATCAACAA 891
QY 200 Ser-----ProSerLeuAlaArgGlnValLeuAsp----- 209
Db 892 TCGAACCCACAGATGAACCAAAATACTATTACGCAAGTAACATCACAAAGTGCACAACTAAA 951
QY 210 -----ThrGlyValAlaValGlnThrTyrSerAlaArgAsnAla 222
Db 952 GAACAGGCATTAAATGTCGCGCGAAACTTAGCTCAAGCTTAAGACAACTGCCAAAAACAAC 1011
QY 223 ValArgThrValLeuAlaProAlaLeuAlaSerArgProAlaValAlaGlnGlyAlaValAsp 242
Db 1012 TTGAATAACTTTAACAATCAATTAAACAATGCACAAAAAGATGCGTTAACGCCGTAGCATTTGAT 1071
QY 243 LeuGlyValSerMetAlaGly-----GlyLeuAlaAlaAsnAlaGlyPheGlyAsn 259
Db 1072 GGTGCAACACACAGTAGCTGTTAAATCAAGAAACTGCAAAAGCAACAGCAATTTAAATAAC 1131
QY 260 ArgLeuLeuSerValGlnSerArgAspHisGlnArgGlyAlaLeuValLeuGlyLeu 279
Db 1132 GCAATGCATAGTTTACAAAT-----GGTATC 1158
QY 280 LysAspLysGluProLysAlaGlnLeuSerGluGluAsnAspTrpLeuGluAlaTyrLys 299
Db 1159 AATGAT-----GAGACACAAACAAACAACTCAGAATACTACCTAGATGCAGAGCCA 1209
QY 300 AlaIleLysSerAlaSerTyrSerGlyAlaAlaLeuAsnAlaGlyLys----- 315
Db 1210 AGTAAGAAATCAGCT-----TATGATCAAGCAGTAATAATGCAGCGAAGCAATTTTAACA 1263
QY 316 ArgMetAlaGlyLeuProLeuAspMetAlaThrAspAlaMetGlyAlaValArgSerLeu 335
Db 1264 AAAGCTAGTGTCAAAATGTAGACAAAGCA-----GCAGTTGAACAAGCAATTGCCAAAT 1317
QY 336 ValSerAlaSerSerLeuThrGlnAsnGlyLeuAlaLeuAlaGlyGlyPheAlaGlyVal 355
Db 1318 GTGAACAGTAGCAAGACGCGGCTTGAACGCTGATGCGAAATTAAATGAAGCTAAAGCAGCT 1377

Db	3409	GGTGCAACACACAGTACCTGGTGTAAATCAAGAAACTGCAGAAAGCAACAGAAATTAATAAC			3468
QY	260	ArgLeuLeuSerValGlnSerArgAspHisGlnArgGlyGlyAlaLeuValLeuGlyLeu			279
Db	3469	GCAATGCATAGTTTACAAAT			3495
QY	280	LysAspLysGluProLysAlaGlnLeuSerGluGluAsnAspTyrPheGluAlaTyrLys			299
Db	3496	AATGAT-----GAGACACAAACAAACAACTCAGAAATACCTAGATGCAGAGCCA			3546
QY	300	AlaIleLysSerAlaSerTyrSerCylAlaAlaLeuAsnAlaGlyLys-----			315
Db	3547	AGTAAGAAATCAGCT-----TATGATCAAGCAGCTAAATGCAGGAAAGCAATTTAACA			3600
QY	316	ArgMetAlaGlyLeuProLeuAspMetAlaThrAspAlaMetGlyAlaValArgSerLeu			335
Db	3601	AAAGCTAGTGGTCAAAATGTAGACAAAGCA-----GCAGTTGAACAAGCATTCGCAAAAT			3654
QY	336	ValSerAlaSerSerLeuThrGlnAsnGlyLeuAlaLeuAlaGlyGlyPheAlaGlyVal			355
Db	3655	GTGAACAGTACGAAGACGGCGTTGACGGTGATGCGAAATTAAATGAAGCTAAAGCAGCT			3714
QY	356	GlyLysLeuGlnGluMetAlaThrLysAsnIleThrAspProAlaThrLysAlaAla---			374
Db	3715	GCGAAA--CAAACGTTAGGTACACTTAACACACATTAATATGACCAACGCTACAGCGTTA			3771
QY	375	-----ValSerGlnLeuThrAsnLeuAlaGlySerAlaAlaValPheAlaGlyTyr			391
Db	3772	GACATGAATTTACACACAGCAACAATGTTGAAGTGTATATACAGTTAAAGCC-----			3825
QY	392	ThrThrAlaAlaLeuThrThrAspProAlaValLysLysAlaGluSerPheIleGlnAsp			411
Db	3826	-----AAAGCGCAACAATTAGATGGTGTATGGGTCAATTAGAAACATCAATTGCTGAT			3879
QY	412	ThrValLysSerThrAlaSerSerThrThrGlyTyr-----			423
Db	3880	-----AAAGACACGACGTTACAAAGTCAAAATTATCAAGATGCTGATGATGCTAAACGA			3933
QY	424	---ValAlaAspGlnThrValLysLeuAlaLysThrVal-----LysAspMetGlyGly			440
Db	3934	ACTGCTTATTCTCAAGCAGTAATGCAGCAGCACTATTTTAAATTAACACAGCTGGCGGT			3993
QY	441	-----GluAlaIleThrHisThrGlyAlaSer			449
Db	3994	AATACACCTAAAGCAGATGTTGAAAGACAAATGCAGCTGTTACACACAGCAAACTACTGCA			4053
QY	450	LeuArgAsnThrValAsn-----			455
Db	4054	TTAAACGCTATTCAAAACTTAGATCGTGCAGAAACAGCGCTGCTAACACAGCGAATTACAAT			4113
QY	456	-----AsnLeuArgGlnArgProAlaArgGluAlaAspIleGluGlyGly			471
Db	4114	GCTTCGGACTTAAATACAAAACAAAAGACATTTAAAGCACACAGTAACAGTGCAGGA			4173
QY	472	ThrAlaAlaSerProSerGluIle			479
Db	4174	CCTGTATCTGCAGCAAAATGGTGT			4197

```

RESULT 13
US-09-287-849-25
; Sequence 25, Application US/09287849
; Patent No. US20020009459A1
; GENERAL INFORMATION:
; APPLICANT: Reed, Steven G.
; APPLICANT: Skeiky, Yasir A.W.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Alderson, Mark
; APPLICANT: Campos-Neto, Antonio
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Fusion Proteins of Mycobacterium tuberculosis Antigens
; TITLE OF INVENTION: and Their Uses
; FILE REFERENCE: 014058-009020US
;

```

```

; CURRENT APPLICATION NUMBER: US/09/287,849
; CURRENT FILING DATE: 1999-04-07
; PRIOR APPLICATION NUMBER: US 08/818,112
; PRIOR FILING DATE: 1997-03-13
; PRIOR APPLICATION NUMBER: US 08/942,578
; PRIOR FILING DATE: 1997-10-01
; PRIOR APPLICATION NUMBER: US 09/025,197
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 09/056,556
; PRIOR FILING DATE: 1998-04-07
; PRIOR APPLICATION NUMBER: US 09/223,040
; PRIOR FILING DATE: 1998-12-30
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 25
; LENGTH: 1797
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:bi-fusion
; OTHER INFORMATION: protein TbH9-Ra35 (designated Mtb59f)
; NAME/KEY: CDS
; LOCATION: (1)..(1791)
US-09-287-849-25

Alignment Scores:
Pred. No.: 0.00229 Length: 1797
Score: 127.50 Matches: 115
Percent Similarity: 33.45% Conservative: 70
Best Local Similarity: 20.80% Mismatches: 179
Query Match: 5.29% Indels: 189
DB: 10 Gaps: 25

US-09-825-414-7 (1-486) x US-09-287-849-25 (1-1797)
QY 20 ThrAlaSerAspAlaSerIleuAlaSerSerSerValArgSerValSerSerAspGln 39
Db 262 ACCGCGGGCAGCGCCGAGCTGACCGCGCCGAGCTCCGGGT----- 303
QY 40 ArgGluIleAsnAlaIleAlaAspTyrIleuThrAspHisValPheAlaAlaHisLys 59
Db 304 -----GCTGCGCGCGCCTACGAGACGGCGGTATGGCTGACG-----GTG 342
QY 60 ProProAlaaspSerAlaaspGlyGlnAla----- 69
Db 343 CCCCCCGCGGTGATCGCCGAGAACCGTGTGACTGATGATTCGTATAGCAGCAACCTC 402
QY 70 -----AlaValAspValHisAsnAlaGlnIleThrAlaIleuIleGlu 83
Db 403 TTGGGGCAAAACACCCCCGGCGATCGCGGTCAACGAGGCCAATACGGCGAGATGTGGGCC 462
QY 84 ThrArgAlaSerArgIleuHisPheGluGlyGluThrProAlaThrIleAlaaspThrPhe 103
Db 463 CAAGACGCGCGCGGATGTTGGCTACGCGCGGCGGACGGCGACGGCGACGGCGACGTTG 522
QY 104 AlaLysAlaGluLysLeuAspArgIleuAlaThrThrThrSerGlyAlaLeu----- 120
Db 523 CTGCCGTTGAGGAGGCGCGGAGATG-----ACCAGCGCGGTGGCTCTCTCGAGCAG 576
QY 121 -----ArgAlaThrProPheAlaMetAlaSer--LeuLeuGlnTyrMet 134
Db 577 GCCCGCCGGGTCGAGAGAGGCTCCGACACCGCGCGCGGAACCAAGTTGATGAACAATGTG 636
QY 135 GlnProAlaIleAsnLysGlyAspTyrPleuProAlaProLeuLysProLeuThrProLeu 154
Db 637 CCCCAGGCGCTGCAACAG-----CTGGCCAGCCCAAGCAGGACACCGCCT--- 684
QY 155 IleSerGlyAlaLeuSerGlyAlaMetAspGlnValGly-----Thr 168
Db 685 ---TCTTCCAAGCTGGGTGCTGTGGAAGACGGTCTCGCCGATCGGTGCGCGATCAGC 741
QY 169 LysMetMetAspArgAlaThrGlyAspLeuHisTyrLeuSerAlaSerProAspArgLeu 188

```


Db 742 AACATGGTGTGCATGGCCCAACACCATGTGCG--ATGACCAACTCGGGTGTGTGCATG 798
QY 189 HisAspAlaMetAlaAlaSerValLysArgHisSerProSerLeuAlaArgGlnValLeu 208
Db 799 ACCAACAACCTTGAGCTGATGTGAAGGCTTTGCTCCGGCGCGCCGCGCAG----- 852
QY 209 AspThrGlyValAlaValGlnThrTyrSerAlaArgAsnAlaValArgThrValLeuAla 228
Db 853 -----GCCGTGCAAGCC---CGCGCGCAAAACGGGGTCCGG----- 885
QY 229 ProAlaLeuAlaSerArgProAlaValGlnGlnGlyAlaValAlaSerMetAla 248
Db 886 ---GCGATGAGCTCGCTGGGCGACGCTCGGTCTTCGGGTCTGGGC----- 930
QY 249 GlyGlyLeuAlaAlaAsnAlaGlyPheGlyAsnArgLeuLeuSerValGlnSerArgAsp 268
Db 931 GGTGGGGTGGCCGCCAACTTGGGT-----CGGGCGGCTCGGTGGTTCG----- 975
QY 269 HisGlnArgGlyGlyAlaLeuValLeuGlyLeuLysAspLysGluProLysAlaGlnLeu 288
Db 976 -----TTG 978
QY 289 SerGluGluAsnAspTyrLeuGluAlaTyrLysAlaIleLysSerAlaSer----- 305
Db 979 TCGGTGCCCGCAGGCGCTGGCGCGGCCCAACCAAGCAGTCAACCCCGCGCGCGCGCTG 1038
QY 306 -----TyrSerGlyAlaAlaLeuAsnAlaGlyLysArgMetAlaGlyLeu 320
Db 1039 CCGCTGACCAGCCTGACCAAGCGCGCGGAAGAAGGGCCCGGCGAGATGCTGGCGGGCTG 1098
QY 321 ProLeuAspMetAlaThrAspAlaMetGlyAlaValArgSerLeuValSerAlaSerSer 340
Db 1099 CCGGTGGGG----- 1107
QY 341 LeuThrGlnAsnGlyLeuAlaLeuAlaGlyGlyPheAlaGlyValGlyLysLeu----- 358
Db 1108 -----CAGATGGCGCCAGGCGCGGTGGGTGAGTGTGCTGCTGCTGCTGCTGCTG 1161
QY 359 -----GlnGluMetAlaThrLysAsnIleThrAspProAla----- 370
Db 1162 CGACCTATGTGATGCCGATCTCTCCGACGCGCGGATATGCCCCCGGCGCTGTGCG 1221
QY 371 -----ThrLysAlaAlaValSerGln 377
Db 1222 CAGGACCGGTTCCGCGACTTCCCGCGCTGCCCTCGACCCGCTCCGCGATGTGCGCCAA 1281
QY 378 Leu-----ThrAsnLeuAlaGlySerAlaAlaValPheAla 389
Db 1282 GTGGGGCCACAGGTGTCAACATCAACACCAACTGGGCTACACACGCGGTGGCGGCC 1341
QY 390 GlyTyrThrThrAlaAlaLeuThrThrAspPro----- 400
Db 1342 GGG-----ACGGGCATCGTCAATCGATCCCAACGGGTGCTGTGACCAACAACAC 1392
QY 401 -----AlaValLysLysAlaGluSerPheIleGlnAspThrValLysSerThrAla 417
Db 1393 GTGATCGCGGGCGCCACCGACATCAATCGCTTCAGCGTGGCTCCGGCCAAACTACGCG 1452
QY 418 SerSerThrThrGlyTyrValAlaAspGlnThrValLysLeuAlaLys----- 433
Db 1453 GTCGATGTGTGGGTATGACCGCACCCAGGATGTGCGGGTGTGACAGTGGCGGTGCC 1512
QY 434 -----ThrValLysAspMetGlyGlyLysAlaIleThrHisThrGlyAlaSer 449
Db 1513 GGTGGCCTGCCCTCGCGCGGAGATCGGTGGCGCGTGGGTGTGAAGCCCGTGTGCGG 1572
QY 450 LeuArgAsnThrValAsnAsnLeuArgGlnArgProAlaArgGluAlaAspIleGluGlu 469
Db 1573 ATGGGCAACAGC----- 1584
QY 470 GlyGlyThrAlaAlaSerProSerGluIleProPheArg 482
Db 1585 GGTGGGCAAGGGCGGAACGCCCGGTGCGGTGCTGCGCAGG 1623

RESULT 14
US-09-860-846-32/c
; Sequence 32, Application US/09860846
; Patent No. US20020164742A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438US1
; CURRENT APPLICATION NUMBER: US/09/860,846
; PRIOR FILING DATE: 2001-05-18
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 32
; LENGTH: 11220
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-860-846-32
Alignment Scores:
Pred. No.: 0.0267 Length: 11220
Score: 127.50 Matches: 113
Percent Similarity: 36.10% Conservative: 61
Best Local Similarity: 23.44% Mismatches: 190
Query Match: 5.29% Indels: 122
DB: Gaps: 18
US-09-825-414-7 (1-486) x US-09-860-846-32 (1-11220)
QY 14 AlaThrAspSerPheArgThrAlaSerAspAlaSerLeu-----AlaSerSerSer 30
Db 8351 GCGGGGAGGCCACAGAGCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 8292
QY 31 ValArgSerValSerSerAspGlnGlnArgGluLeuAsnAlaIleAlaAspTyrLeuThr 50
Db 8291 ACCAGTCCGTAAAGTCCGGAGGAGACAGCACTCGGTGCCACCGCGCGCGCGCGCG 8232
QY 51 AspHisValPheAlaAlaHisLysLeuProProAlaAspSerAlaAspGlyGlnAlaAla 70
Db 8231 GCGAGACCGTCCGGCTG-----CGGCCAGCAGACCGCGCGCGCGCGCGCGCTGC 8181
QY 71 Val-----AspValHisAsnAlaGlnIleThrAlaLeuIleGluThrArgAla 86
Db 8180 GCGGCACCGTCCCAAGCGGCTCCACTCCAGCAGGTGCAGCGCGGCTCCAGAGTCCGG 8121
QY 87 SerArgLeuHisPheGluGlyGluThrProAlaThrIleAlaAspThrPheAlaLysAla 106
Db 8120 AAGGCGCCAGCTGCGGGGGTGCAGGGGAGCAGCAGCAGTGAAGGAGTCCGGCGGAAC 8061
QY 107 -----GluLysLeuAspArgLeuAlaThrThrThrSerGlyAlaLeuArgAlaThr 123
Db 8060 GGTGCCCGGAGAGTCCGGCGGCTCAGCAGACCGGTGTGCGGGCGCGCGCGCGCGAG- 8002
QY 124 ProPheAlaMetAlaSerLeuLeuGlnTyrMetGln-----ProAlaIleAsnLysGly 141
Db 8001 CCGGACCGGAGGGCGGTGCGCGCCGACCGCGCTACAGAGAGATCCCGCTCCAGCGGAAC 7942
QY 142 AspTyrLeuProAlaProLeuLysProLeuThrProLeuIleSerGlyAlaLeuSerGly 161
Db 7941 AGCGCGGTGCCCGCGCGGAGACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 7891
QY 162 AlaMetAspGlnValGlyThrLysMetMetAspArgAla-----ThrGlyAspLeu--- 178
Db 7890 GCGGTGAGCAGCGCGCGGTGAAGCGCGGAGAACCGCGCGCGCGCGCGCGCGCGCGCT 7831
QY 179 -----HisTyrLeuSerAlaSerProAspArgLeuHisAspAlaMet 192
Db 7830 CCGGAGGCGCAGTCCGGGAGAACCTGTGCGCAGCGCGCGCGCGCGCGCGCGCGCGCTG 7771

QY 193 -----AlaAlaSerValLysArgHisSerProSerLeuAlaArgGlnValLeuAsp 209
Db 7770 GAAGAGGGGGCCGTAGCCGTA-----GCCGTTCCGGCGAAGCGCTCGTA 7726
QY 210 ThrGlyValAlaValGlnThrTyrSerAlaArgAsnAlaValArgThrValLeuAlaPro 229
Db 7725 CAGACCGTCCACGTCCACCGGCTCGGCGCCCGCGG----- 7690
QY 230 AlaLeuAlaSerArgProAlaValGln-----GlyAlaValAspLeuGly 244
Db 7689 -----CGGCCAGCCCTCCGGGTCCGGCGAGGGGGCGGTCCGGTCCGACGGGC 7642
QY 245 ValSerMetAlaGlyGlyLeuAlaAlaAsnAlaGlyPheGlyAsnArgLeuLeuSerVal 264
Db 7641 GGCCAGCACACCGGTGGCGTGCCTCCACTCCGCTCGCCCGCGGTCTCCGGGGTG 7582
QY 265 GlnSerArgAspHisGlnArgGlyGlyAlaLeuVal----- 276
Db 7581 CGCGTAGAGCCCGAAGGTACAGACGCCGGAAGTCTGCTCGCGCCGACGAGACAGTGCAC 7522
QY 277 -----LeuGlyLeuLysAspLysGlnProLysAlaGlnLeuSerGlnGlu 291
Db 7521 ACGGACCGCGCCACGACGAGGGGACGACGAGCGCGGCTCGAGGGTGAGCTCTCGACACAG 7462
QY 292 -----AsnAsp-TyrLeuGlnAlaTyrLysAlaIleLysSerAlaSerTyrSerGlyAl 309
Db 7461 ATCGCAACCGACCTGTGTCCTCCCGGCTCGGAACCGCACGTCCACGAAACCGCTTCCCGGACG 7402
QY 309 aAlaLeuAsnAlaGly----- 314
Db 7401 CAGCAC-GGTGCCGCGCCACCGCTGTGTCGCCGACGAGGGGTGCGTAGCGAGGAGAGGC 7343
QY 315 -----LysArgMe 317
Db 7342 TCCCGGTGAGCAGGACGCGTCCGAGTCCGCGAGCGCCACGCGCGCGAGCAGCGGGT 7283
QY 317 tAlaGlyLeuProLeuAspMetAlaThrAspAlaMetGlyAlaValArgSerLeuValSe 337
Db 7282 GCTCGCGCGCCCGAGACCGCGGAGGTGATGTACCGCGCGCGG-AGAGTCCGGGCTGC 7224
QY 337 rAlaSerSerLeuThrGlnAsnGlyLeuAlaLeuAla-----GlyGlyPheAlaGlyVa 355
Db 7223 GGCCAGTAGCGCTCGTCTGGAAGCGGTAGGTGGGACATCGGGGTGGTGTCTG-CT 7165
QY 355 lGlyLysLeuGlnGlnMetAlaThr----- 363
Db 7164 GGGGAGGAGAGAGGCCCACTGACGCGTGAGCGCCGTGGCCACGAGGCTCGGCGAGGAGGT 7105
QY 364 ----LysAsnIleThrAspProAlaThrLysAlaAlaValSerGlnLeuThrAsnLeuAl 382
Db 7104 GGTGAGCGGTGCTGTCCGCGCTGTGTACGCGCGGAGGTTGCCGAGTCCGGTGCAGGCTTC 7045
QY 382 aglySerAlaAlaValPheAlaGlyTyrThrThrAlaAlaLeuThrThrAspProAlaVa 402
Db 7044 GGGCAGCGCCATGTGTGAGACGGGGTG-----GCGCTGACCTCGACGAAGTGGGT 6994
QY 402 lLys 403
Db 6993 GAAG 6990

RESULT 15
US-09-861-289-32/c
; Sequence 32, Application US/09861289
; Patent No. US20020110897A1
; GENERAL INFORMATION:
; APPLICANT: Sherman, D.H.
; APPLICANT: Liu, H.
; APPLICANT: Xue, Y.
; APPLICANT: Zhao, L.
; TITLE OF INVENTION: DNA encoding methymycin and pikromycin
; FILE REFERENCE: 600.438051
; CURRENT APPLICATION NUMBER: US/09/861,289

; CURRENT FILING DATE: 2001-05-18
; PRIOR APPLICATION NUMBER: 09/105,537
; PRIOR FILING DATE: 1998-06-26
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 32
; LENGTH: 11220
; TYPE: DNA
; ORGANISM: Streptomyces venezuelae
US-09-861-289-32

Alignment Scores: 0.0267 Length: 11220
Pred. No.: 127.50 Matches: 113
Score: 36.10% Conservative: 61
Percent Similarity: 23.44% Mismatches: 190
Best Local Similarity: 5.29% Indels: 122
Query Match: 10 Gaps: 18
DB:

US-09-825-414-7 (1-486) x US-09-861-289-32 (1-11220)

QY 14 AlaThrAspSerPheArgThrAlaSerAspAlaSerLeu-----AlaSerSerSer 30
Db 8351 GCGGGCAGGCGCCACGACGAGCGGTCCGCGGGGCGGCTCTCGCCCCGCTCGACGGCTCC 8292
QY 31 ValArgSerValSerSerAspGlnArgGlyLysAlaIleAlaAspTyrLeuThr 50
Db 8291 ACCAGGTCCGTAAAGTCCGGGGAAGACAGGACCTCTGCCACCGGCGCGACCGCGCG 8232
QY 51 AspHisValPheAlaAlaHisLysLeuProProAlaAspSerAlaAspGlyGlnAlaAla 70
Db 8231 GCGAGACCGTCCGCGCTCG-----CCGCCACGACGACCGCGCGCGGCGAGGCGCTGC 8181
QY 71 Val-----AspValHisAsnAlaGlnIleThrAlaLeuIleGlyThrArgAla 86
Db 8180 GCGGACACGTCCACGCGGCTCCACTCCAGCAGGTGCAGCGGTCCAGAGTCCGGGTCTG 8121
QY 87 SerArgLeuHisPheGlnGlyGlyThrProAlaThrIleAlaAspThrPheAlaLysAla 106
Db 8120 AAGCGCGCCAGCTCGCGGGGTCCGACGCGGACGACCGGTGAGGAGTCCGGCGGGAACAC 8061
QY 107 -----GluLysLeuAspArgLeuAlaThrThrSerGlyAlaLeuArgAlaThr 123
Db 8060 GGCTGCCCGGAGGAGTCCGCGCGGCTCACGAGACACCTGTCCGGGCGGCGGCGCCAG- 8002
QY 124 ProPheAlaMetAlaSerLeuLeuGlnTyrMetGln-----ProAlaIleAsnLysGly 141
Db 8001 CCGCACGCGGAGGGCGGTGGCGCCGACCGCGGTACAGAGATCCCGCTCCAGGCGAAGCGG 7942
QY 142 AspTyrLeuProAlaProLeuLysProLeuThrProLeuIleSerGlyAlaLeuSerGly 161
Db 7941 AGCCCGTGCCTCCGCGCGAAGCGCCCGCGCGC-----ACCGCGCGCTGCACGGC 7891
QY 162 AlaMetAspGlnValGlyThrLysMetMetAspArgAla-----ThrGlyAspLeu--- 178
Db 7890 GCGGTGAGCAGCGCGGGGTGAAGCGCGAAGCGCGCGCTCGGACCGCGGCACTCGGC 7831
QY 179 -----HisTyrLeuSerAlaSerProAspArgLeuHisAspAlaMet 192
Db 7830 CCGCAGGGCCACGTCCGCGAAGACCTCTCGCCACCGCCGACAGACACGAGCGCCCTG 7771
QY 193 -----AlaAlaSerValLysArgHisSerProSerLeuAlaArgGlnValLeuAsp 209
Db 7770 GAAGAGGGGGCGGTAGCCGTA-----GCCGTTCCGGCGAAGCGCTCGTA 7726
QY 210 ThrGlyValAlaValGlnThrTyrSerAlaArgAsnAlaValArgThrValLeuAlaPro 229
Db 7725 CAGACCGTCCACGTCCACCGGCTCGGCGCCCGCGG----- 7690
QY 230 AlaLeuAlaSerArgProAlaValGln-----GlyAlaValAspLeuGly 244
Db 7689 -----CGGCCAGCCCTCCGGGTCCGGCGAGGGGGCGGTGCGGTCCGACCGGCGG 7642

QY	245	ValSerMetAlaGlyLeuAlaAlaAsnAlaGlyPheGlyAsnArgLeuLeuSerVal	264
		:::	:::
Db	7641	GGCCAGCACACCGGTGGCGTCCGCTCCGCCCTCGCCCCGGCGCTCTCCGGGTG	7582
QY	265	GlnSerArgAspHisGlnArgGlyGlyAlaLeuVal-----	276
		:::	
Db	7581	CGCGTAGAGCCCGAAGGTACGACGCCCGGACTCGTCCGCTCGCGCCGACGACAGCTGCAC	7522
QY	277	-----LeuGlyLeuLysAspLysGluProLysAlaGlnLeuSerGluGlu	291
		:::	:::
Db	7521	ACGGACCGCCGACGACGGGGCAGCACGAGCGCGCTCGAGGGGTGAGTCTCTCGACACAG	7462
QY	292	-----AsnAsp-TrpLeuGluAlaTyrLysAlaIleLysSerAlaSerTyrSerGlyAl	309
		:::	:::
Db	7461	ATCGCAACCGACTGTGCTCCCGCTCGGAACGCTCACCAGAACGCCGTCTCCGGGACAG	7402
QY	309	aAlaLeuAsnAlaGly-----	314
		:::	
Db	7401	CAGCAC-GGTGCCGGCCACCGCGTGTCCGCCAGCCAGGGGTGCGTACGAGAGGAGAGGC	7343
QY	315	-----LysArgMe	317
		:::	:::
Db	7342	TCCCGGTGAGCAGCAGCCGCTCGAGTCCCGGAGCGCCACGCGCCGCGGACAGCGGGT	7283
QY	317	tAlaGlyLeuProLeuAspMetAlaThrAspAlaMetGlyAlaValArgSerLeuValSe	337
		: :::	:::
Db	7282	GCTCGGCCCGCCGAGACCGCGGAGGTGATGTCAACGGCGGGCGG-AGAGTCCGGGCTGC	7224
QY	337	rAlaSerSerLeuThrGlnAsnGlyLeuAlaLeuAla-----GlyGlyPheAlaGlyVa	355
		::: :::	:::
Db	7223	GGCCAGTAGCGCTCGGTCTGGAAGGCCGTAGGTGGGCAATCGGGGTGGTGTCTGTG-GT	7165
QY	355	lGlyLysLeuGlnGluMetAlaThr-----	363
		:::	:::
Db	7164	GGGAGGAGAGAGGCCCACTCGACGCTGAGGCCGTTGGCCACAGGCTCGGCGAGGAGGT	7105
QY	364	---LysAsnIleThrAspProAlaThrLysAlaAlaValSerGlnLeuThrAsnLeuAl	382
		:::	:::
Db	7104	GGTGAGCGCGTGTCTCCCGCTGTGTACAGCGGAGGCTGCCGAGTCCGGTGACGGTCTC	7045
QY	382	aglySerAlaAlaValPheAlaGlyTyrThrThrAlaAlaLeuThrThrAspProAlaVa	402
		: :::	:::
Db	7044	GGGCAGCGCCATGTGTAGAGACGGGGTGG-----GCCGTGACCTCGACGAAGTGGGT	6994
QY	402	Lys	403
Db	6993	GAAG	6990

Search completed: January 31, 2003, 07:21:46
Job time : 138.921 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: January 31, 2003, 02:25:34 ; Search time 3127.71 Seconds
(without alignments)
4531.450 Million cell updates/sec

Title: US-09-825-414-66
Perfect score: 2407
Sequence: 1 MHINSAQQPGVAMESFRT.....EEGISAFSRSETPFQLRRL 487

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2054640 segs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DEV=xlp
-O=/cgn2_1/USPTO_spool/US09825414/runat_28012003_161741_26085/app.query.fasta_1.1294
-DB=GenEmbl -QFMT=fastap -SUFMT=p2n.rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=biosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09825414@cgn_1_1_4707@runat_28012003_161741_26085 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*

29: em_vl:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	1869.5	77.7	52498	1	AF232004	AF232004 Pseudomon
2	1591.5	66.1	1834	1	AF458051	AF458051 Pseudomon
3	460	19.1	208050	1	AL646083	AL646083 Ralstonia
C 4	160.5	6.7	1558	3	AY094894	AY094894 Drosophila
C 5	160.5	6.7	163709	3	AC108877	AC108877 Drosophila
C 6	150.5	6.3	203050	1	AL646078	AL646078 Ralstonia
C 7	149.5	6.2	10029	1	AE008061	AE008061 Agrobacte
C 8	149.5	6.2	12020	1	AE009096	AE009096 Agrobacte
9	149	6.2	12313	1	AE008285	AE008285 Agrobacte
C 10	149	6.2	12668	1	AE009328	AE009328 Agrobacte
C 11	146.5	6.1	301769	3	AE003482	AE003482 Drosophila
C 12	146	6.1	306250	1	SME591788	AL591788 Sinorhizo
C 13	145	6.0	13040	1	AE005769	AE005769 Caulobact
C 14	143.5	6.0	6413	1	AF067776	AF067776 Abiotroph
C 15	143	5.9	132457	9	AC016584	AC016584 Homo sapi
C 16	142	5.9	16675	1	AE011814	AE011814 Xanthomon
C 17	142	5.9	138685	2	AC017582	AC017582 Drosophila
C 18	142	5.9	166249	3	AC093502	AC093502 Drosophila
19	142	5.9	185404	3	AC104703	AC104703 Drosophila
C 20	142	5.9	341319	3	AE003536	AE003536 Drosophila
21	141	5.9	9083	1	AE008286	AE008286 Agrobacte
C 22	141	5.9	11449	1	AE005078	AE005078 Halobacte
C 23	140.5	5.8	291150	1	AP003135	AP003135 Staphyloc
C 24	140.5	5.8	342600	1	AP003363	AP003363 Staphyloc
25	140	5.8	1512	1	PER297534	AJ297534 Pseudomon
26	140	5.8	7100	6	AX004713	AX004713 Sequence
C 27	140	5.8	134416	7	AF020713	AF020713 Bacteriop
28	140	5.8	213680	1	BSUB0012	Z99115 Bacillus su
29	139.5	5.8	1473	1	AB073918	AB073918 Campyloba
30	139.5	5.8	8923	1	AB015223	AB015223 Staphyloc
C 31	139.5	5.8	304050	1	AP004829	AP004829 Staphyloc
32	138.5	5.8	1473	1	AB080202	AB080202 Campyloba
33	138.5	5.8	3433	1	SDY243529	AJ243529 Streptoco
34	138.5	5.8	12312	1	AE012307	AE012307 Xanthomon
35	138	5.7	3831	6	AX066471	AX066471 Sequence
36	138	5.7	6136	6	AX066467	AX066467 Sequence
37	138	5.7	8979	6	AX122823	AX122823 Sequence
38	138	5.7	10635	1	AE011815	AE011815 Xanthomon
39	138	5.7	13431	1	AE000234	AE000234 Escherich
40	138	5.7	16902	1	D90775	D90775 E.coli geno
41	138	5.7	18700	1	D90774	D90774 E.coli geno
42	138	5.7	47233	2	AC017267	AC017267 Drosophila
C 43	138	5.7	156251	2	AC107325	AC107325 Drosophila
C 44	138	5.7	194657	2	AC009376	AC009376 Drosophila
45	138	5.7	287649	3	AE003516	AE003516 Drosophila

RESULT 1

ALIGNMENTS

AF232004/c	AF232004	52498 bp	DNA	linear	BCT 05-MAR-2001
LOCUS					
DEFINITION	Pseudomonas syringae pv. tomato strain DC3000 Hrp pathogenicity island, complete sequence.				
ACCESSION	AF232004 AF061028 AF061029 AF232006 L41861				
VERSION	AF232004.3 GI:13325077				
KEYWORDS	.				
SOURCE	Pseudomonas syringae pv. tomato.				
ORGANISM	Pseudomonas syringae pv. tomato Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae; Pseudomonas.				
REFERENCE	1 (bases 25494 to 29778)				
AUTHORS	Preston,G., Huang,H.C., He,S.Y. and Collmer,A.				
TITLE	The HrpZ proteins of Pseudomonas syringae pvs. syringae, glycinea, and tomato are encoded by an operon containing Yersinia ysc homologs and elicit the hypersensitive response in tomato but not soybean				
JOURNAL	Mol. Plant Microbe Interact. 8 (5), 717-732 (1995)				
MEDLINE	96025089				
PUBMED	7579616				
REFERENCE	2 (bases 22134 to 25847; 29687 to 32670)				
AUTHORS	Deng,W.L., Preston,G., Collmer,A., Chang,C.J. and Huang,H.C.				
TITLE	Characterization of the hrpC and hrpRS operons of Pseudomonas syringae pathovars syringae, tomato, and glycinea and analysis of the ability of hrpF, hrpG, hrpC, hrpT, and hrpY mutants to elicit the hypersensitive response and disease in plants				
JOURNAL	J. Bacteriol. 180 (17), 4523-4531 (1998)				
MEDLINE	98389667				
PUBMED	9721291				
REFERENCE	3 (bases 31672 to 51723)				
AUTHORS	Charkowski,A.O., Alfano,J.R., Preston,G., Yuan,J., He,S.Y. and Collmer,A.				
TITLE	The Pseudomonas syringae pv. tomato HrpW protein has domains similar to harpins and peptate lyases and can elicit the plant hypersensitive response and bind to peptate				
JOURNAL	J. Bacteriol. 180 (19), 5211-5217 (1998)				
MEDLINE	98422476				
PUBMED	9748456				
REFERENCE	4 (bases 901 to 22404; 31672 to 51723)				
AUTHORS	Alfano,J.R., Charkowski,A.O., Deng,W.L., Badel,J.L., Petnicki-Ocwieja,T., van Dijk,K. and Collmer,A.				
TITLE	The Pseudomonas syringae Hrp pathogenicity island has a tripartite mosaic structure composed of a cluster of type III secretion genes bounded by exchangeable effector and conserved effector loci that contribute to parasitic fitness and pathogenicity in plants				
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 97 (9), 4856-4861 (2000)				
MEDLINE	20243785				
PUBMED	10781092				
REFERENCE	5 (bases 1 to 52498)				
AUTHORS	Ramos,A.R., Rehm,A.H. and Collmer,A.R.				
TITLE	Pseudomonas syringae pv. tomato DC3000 hrpL through hrpU unpublished				
JOURNAL	6 (bases 1 to 52498)				
AUTHORS	Alfano,J.R. and Collmer,A.				
TITLE	Direct Submission				
JOURNAL	Submitted (07-FEB-2000) Dept. Biol. Sci., UNLV, 1854 Maryland Parkway, Las Vegas, NV 89154, USA				
REFERENCE	7 (bases 1 to 52498)				
AUTHORS	Ramos,A.R., Rehm,A.H. and Collmer,A.R.				
TITLE	Direct Submission				
JOURNAL	Submitted (22-NOV-2000) Plant Pathology, Cornell University, 334 Plant Sciences Bldg., Ithaca, NY 14850, USA				
REMARK	Sequence update by submitter				
REFERENCE	8 (bases 1 to 52498)				
AUTHORS	Ramos,A.R., Rehm,A.H. and Collmer,A.R.				
TITLE	Direct Submission				
JOURNAL	Submitted (05-MAR-2001) Plant Pathology, Cornell University, 334 Plant Sciences Bldg., Ithaca, NY 14850, USA				
REMARK	Sequence update by submitter				
COMMENT	On or before Mar 14, 2001 this sequence version replaced gi:3228544, gi:790906, gi:3228541, gi:8037790, gi:11276506.				
FEATURES	location/Qualifiers				
SOURCE	1. .52498				

/organism="Pseudomonas syringae pv. tomato"
 /strain="DC3000"
 /db_xref="taxon:323"
 /note="exchangeable effector locus flanking the hrp/hrc cluster
 pathovar: tomato"
 complement(908. .1972)
 /gene="queA"
 complement(908. .1972)
 /gene="queA"
 /codon_start=1
 /transl_table=11
 /product="Quea"
 /protein_id="AAf71483.1"
 /db_xref="GI:8037772"
 /translation="MRVADFFELPDSLIARHPLAERSSRLTLDPGALAHROFT
 DLEHRSGLDMVFNNTRVLPARLFGQKASGKLELLVERVLDSHRVLAHVASKSPK
 PGSSILIDGGGEAEMVARHDALELFEAEVLPLIDRVGHMLPPIIDRDEGADRER
 YQYVAQRAGAVAAPTAGLHPDQPLMEAIAAKGVETAFVTLHVAGATFQYRVEQIED
 HHMSEMLEVSDVDVAACRARGRVAVGTSYRSLSAARDGQKPESGDTDIF
 IYPRFPHVDALVTNFEHPSTLLMLVSAFAGYPTMAAYAAIEHGYPFSYGDAM
 FITRNPAPTAPQESAPEDHA"
 2090. .2177
 /product="trRNA-Leu"
 2343. .2753
 /note="ORF4"
 /codon_start=1
 /transl_table=11
 /product="unknown"
 /protein_id="AAf71484.1"
 /db_xref="GI:8037773"
 /translation="MNKIVYKAYFKPIGEEVSVKVPTGEIKKGFPGDEIMKETQW
 QQTGWSDCQIDGERLSKDVESAVALQINADGEIQYVLPILSGAYDALKYREIRHNR
 TELSPGQSYVEGYGSFTEGTVLVAKKFQSSAS"
 complement(2908. .3324)
 /note="ORF3"
 /codon_start=1
 /transl_table=11
 /product="unknown"
 /protein_id="AAf71485.1"
 /db_xref="GI:8037774"
 /translation="MRAYKNLTAKIGGFLALTTIGTSLPAFVNDCLDNDNSTGAT
 CGGNDKDLNDNVTDAAFGNDKMDNDHHTDAFGNDKDLNDHHTDAFGNDKDL
 LDNDNKTDAAFGGNDRDLNDNNTDNYNGTPSAKK"
 complement(3939. .4664)
 /note="ORF2"
 /codon_start=1
 /transl_table=11
 /product="unknown"
 /protein_id="AAf71486.1"
 /db_xref="GI:8037775"
 /translation="MOKTTLNKLAFAMLAGCVSGPAGSDIOGAQAEKTPVKLNLD
 AYTSKRIDAVLEARTNKSYNKQGLIDLVSGAFGLGTPYRSNMLVGSANVPEQLYDFR
 GLDCFYLDYVEAFRRSTQODEVRNLVQRYKGGVDLELRKHFTDNAVGTAYPVA
 DDITAQISPGAVSVRKRLNERAKGKVYLPGLPYVERSWTYIPSRVLVDSQVYVSHLRTGD
 YIGIYTPASRAGCDTRRFLYRDG"
 5571. .6410
 /gene="tnpA"
 5571. .6410
 /gene="tnpA"
 /note="truncated at the amino terminus"
 /codon_start=1
 /transl_table=11
 /product="tnpA"
 /protein_id="AAf71487.1"
 /db_xref="GI:8037776"
 /translation="MGFLRPAYPLMAMLRITHLMQNMFDYSDPAMEALYETTLRQFA
 GLSLRIADETTILNFRRLLEKHELAGGILQVINGLGDRLMLROGMVVDAITIIHAP
 SSTKNKDGKRDPEMHQTKKNDYFFGMAKHIGVDAESGLVHSLVGTAAVADVTOVDQ
 ILSHEETVYVSGDAGYTGVDKRAEHQDROMIWSIARPSRYKHGEKSLIARYRKIEE
 TKAQLEAKVEHPLRVYIKRQFGYTKVFRGLAKNTAQOATLFALSNLMVMYKRLLAMGE
 VRL"
 complement(6656. .8056)


```

/Note="ORF1"
/codon_start=1
/transl_table=11
/product="unknown"
/protein_id="AAF71488.1"
/db_xref="GI:8037777"
/translation="MRPVGGPAPGYPPPTYEAERPTQAAGNDRARSSQASSSPASV
APETPMGLDLKREFPAGRYPDMKYNIRLKEGOEPGGKDYVHTRRRKPDAAGSSHVH
GGQSVASTASAQSKALQDTNFRASDLAELARWCESHPYALAPSKAAGKSSQLSAPV
VSILOEGKHALEORLEAQLADYVSEGRDLHLINLNTLEMDSCLGTSGKLWAPD
SNDKLIATAKARYFDDEFNAQLPELAPLTMKMSKDSLGMRELLRDAPGLVIGEGHNS
TSSKRELIINMKSILKASGVTTLEMEHLCAESHDKALNNYLSAPKGSMPARLKNYLDL
OSQGHQAPBELHTKYNFTTLVEAKHAGLRVSLDTSITYAPDEKAEIKRAQAMNYA
AEKIRLSKEGKWAFVGGATHATSCDGVPGLAELHGVRSVLDLGLKSRATVDINVK
NYGKLNPDVRLSYKV"
complement(8131. .10473)
/gene="hrpk"
complement(8131. .10473)
/gene="hrpk"
/codon_start=1
/transl_table=11
/product="Hrpk"
/protein_id="AAF71489.1"
/db_xref="GI:8037778"
/translation="MRISSEPEVIYNQPTPELALAVESPLAKALPTPVGGGQAGVQ
EGQPAQNTQGAPTGAEOQTASSILALLQSGGAANANVNVQPPQPCADSPAEAPPT
EATAPAPVEQVAAPTNTDATDASNAKADAFLDNSEYSSPEALKRWEPVAHLPP
EEREQAAKELNRPITAAWMAWDGPDAGKAMDFINANPALKAVDVAQGGGNADKIT
NKDLKTFKAMMEKADADKDLANTKDNPDADPQSLNEMVRSAAVMRANMPLAIADP
HHAVDADTKVDGNVDAEDLKQSNPGLSKALKQSCSTWQPGFLQGVDEAGMSG
RKKAAHSPDKMFDKAKNLSEWIKKSAPTINGQFASMLSDSATINAVAGIDISKLDKVF
DKPKSYSGAQKAAVMVKLOQTQOSVYAGRSLRNTDKTEQGLNDRISLOADPDVQAYL
NKSIPQERNLVRSASLQKAVVEOTKVNNSGQALQDTMDKADKAVNKNHPNADYSGA
ISGLSAQLOLQKDLFPDSKVPPTNOVLENKPDLDKIATSVYTNSEGGALKQLGOK
KSDASQSLQTDNQKAIYDSVLPDPFTENQHENYMNSTLGTQDSKKRKLLEGKSD
GGPSMLTQLAEGVGGKALNSIVGRASVSSLIAHGDKLGATQAIYDSTRMGAELKGG
IDAGAKMWGREASAGLRLGGQMIGRAVMVAGEATGLAAGALGAALPVVGWAIDGA
LAVFGISMIIDAVKHKHAKQAFDHNVDVLDQFGIAKAH"
complement(10560. .11114)
/gene="hrpl"
complement(10560. .11114)
/gene="hrpl"
/Note="alternative sigma factor"
/codon_start=1
/transl_table=11
/product="Hrpl"
/protein_id="AAG33875.2"
/db_xref="GI:13325078"
/translation="MFQKIVILDSIQPRQSSSAGIRQMTADQIQMLRAFIQKRVMP
DDVDILQCVLEALRNEHKFQMSKPDQWLGCIALNLIRNHRKMYROPYQESWEDE
VHSELEGHGDVSHOVEGHRQLARIYQIADCLPSNMQKVLEYSLEMGNYQETANSIGV
PIGTVRSRLSGARVOLKQIDPFA"
11358. .12464
/gene="hrpj"
11358. .12464
CDS
Alignment Scores:
Pred. No.: 3,21e-106 Length: 52498
Score: 1869.50 Matches: 381
Percent Similarity: 85.19% Conservative: 33
Best Local Similarity: 78.40% Mismatches: 71
Query Match: 1 Indels: 1
Gaps: 1
US-09-825-414-66 (1-487) x AF232004 (1-52498)
QY 1 MethisileasnglnSeralaglnInProcglyValAlametGluSerpheargThr 20
DB 46117 ATGCACATCAACCGCGCTGCCAACACCGCTGTGACTGCGACGAGATAGCTTTCGGACA 46058
QY 21 AlaSerAspAlaSerLeuAlaSerSerValargSerValserThrThrSerCysarg 40
DB 46057 GCGTCCGACGCGCTCTTGCTCCAGCTCTGCGATCTGTCAAGCTCCGATCAGCAACGC 45998
```

```

QY 41 AspleuglnAlaIleThrAspTyrLeuLysHisValPheAlaAlaHisargPheSer 60
DB 45997 GAGATAAATGCGATTGCCGATTACCTGACAGATCATGTGTTGCTGCGCATAACTGCCG 45938
QY 61 ValIleGlySerProaspGluargaspAlaAlaLeuAlaHisasnGluGlnIleaspAla 80
DB 45937 CCGGCCGATTGCGTGTATGCCCAAGCTGCAGTTGACGTTACACAATAGCGCAGATCCTGCG 45878
QY 81 LeuValGluThrargAlaasnArgLeuTyrSerGluGlyGluThrProAlaThrIleAla 100
DB 45877 CTGATCGAGACGCGCCGACAGCCGCTGCACCTGCAAGGGGAACCCGGCAACCATCGCC 45818
QY 101 GluThrPheAlaLysAlaGluLysPheaspArgLeuAlaThrThrAlaSerSeralPhe 120
DB 45817 GACACCTCGCCCAAGCGGAAAGCTCGACCGATTGGCCGACGACTACATCAGCGCGCTTG 45758
QY 121 GluAsnThrProPheAlaAlaAlaSerValLeuGlnItyrMetGlnProAlaIleasnLys 140
DB 45757 CGGGCGACGCGCTTTCCTCATGGCTCGTTCAGTACATGACAGCTCGCATCAACAAG 45698
QY 141 GlyAspTrpLeuAlaThrProLeuLysProLeuThrProLeuIleSerGlyAlaLeuSer 160
DB 45697 GCGGATTGGCTGCGGCTCCGCTCAAAACCGCTGACCCCGCTCATTTCCGGACCGCTGTCG 45638
QY 161 GlyAlaMetaspGlnValGlyThrLysMetMetaspArgAlaargGlyaspLeuHisTyr 180
DB 45637 GCGCGCATGACACAGTGGGCACCAAGATGATGACCCGCGACGCGGTGATCTGCAATTAC 45578
QY 181 LeuSerThrSerProaspLysLeuHisAspAlaMetAlaValSerValysargHisSer 200
DB 45577 CTGAGCGCCTCGCCGACAGGCTCCACGATGCGATGGCCGCTTCGTTGAAGCCCACTCG 45518
QY 201 ProAlaLeuGlyArgGlnValValaspMetGlyIleAlaValGlnThrPheSerAlaLeu 220
DB 45517 CCAAGCCTTGCTCGACAGGTTCTGGACACGGGGTTCGGGTTCAAGACGTACTCGCGCCGC 45458
QY 221 AsnValValargThrValLeuAlaProAlaLeuAlaSerargProserValGlnGlyAla 240
DB 45457 AACGCCGTACGTACCGTATTGGCTCCGGCACTGGCGTCCAGACCCGCGTGCAGGGTGCT 45398
QY 241 ValAspPheGlyValSerThrAlaGlyGlyLeuValAlaasnAlaGlyPheGlyaspArg 260
DB 45397 GTGGACCTTGCTGTATCGATGGCGGGGTGTGCTGCTGCCAACGACGAGCTTGGCAACCGC 45338
QY 261 MetLeuSerValGlnSerargaspGlnLeuArgGlyGlyAlaPheValLeuGlyMetLys 280
DB 45337 CTGCTCAGTGTGACGTGCGGTGATCACACAGCGTGGCGGTGCATTAGTGTGCTTGAAG 45278
QY 281 AspLysGluProLysAlaAlaLeuSerGluGluThrAspTrpLeuAspAlaTyrLysAla 300
DB 45277 GATAAAGAGCCCAAGGCTCACTGAGCGAAGAAACGACTGGCTCGAGGCTTATAAAGCA 45218
QY 301 IleLysSerAlaSerTyrSerGlyAlaAlaLeuAsnAlaGlyLysArgMetAlaGlyLeu 320
DB 45217 ATCAATCGGCCAGCTACTCGGGTGGCGGCTCAACGCTGGCAAGCGGATGCCGCTCTG 45158
QY 321 ProLeuAspValAlaThraspGlyLeuLysAlaValargSerLeuValSerAlaThrSer 340
DB 45157 CCACTGATATGCGACGACGCAATGCGTGGCTGAAGACCTGTGTCAGCTGCCAGC 45098
QY 341 LeuThrLysasnGlyLeuAlaLeuAlaGlyGlyTyrAlaGlyValSerLysLeuGlnLys 360
DB 45097 CTGACCCAAAACGGTCTGGCCCTGCGGGGTGCTTTCAGAGGGTAGGCAAGTTGCAGAGAG 45038
QY 361 MetAlaThrLysasnIleThraspSerAlaThrLysAlaAlaValSerGlnLeuSerAsn 380
DB 45037 ATGGCGACGAAATAATATCACCGACCCGCGCACCAAGCGCGGTACAGTTCAGCAAC 44978
QY 381 LeuValGlySerValGlyValPheAlaGlyTyrThrThrAlaGlyLeuAlaThrAspPro 400
DB 44977 CTGGCAGGTTGCGCAGCGGTTTTCGACAGGCTGGACCAAGCGCGGCTGACCAACGATCCC 44918
```

QY 401 AlavalysLysAlaGluSerPheIleGlnAspLysValLysSerThrAlaSerThr 420
Db 44917 GCGGTGAAAAAAGCCGAGTCGTTTCATACGACACCGGTGAATCGACTCATCCAGTACC 44858
QY 421 ThrSerTyrValAlaAspGlnThrValLysLeuAlaLysThrValLysAspMetSerGly 440
Db 44857 ACAGGCTACGTAGCCGACACGACCGTCAACTGGCGAAGACCGTCAAGACATGGCGGG 44798
QY 441 GluAlaIleSerSerThrGlyAlaSerIleuArgSerThrValAsnAsnLeuArgHisArg 460
Db 44797 GAGCGGATCACCACATACCGGCCAGCTGGCCCAATACGTCATACCTGCGTCAACGC 44738
QY 461 SerAlaProGluAlaAspIleGluGlyGlyIleSerAlaPheSerArgSerGluThr 480
Db 44737 CCGGCTCGTGAAGCTGATATAGAGAGGGGGCGACGGCGCT--TCTCCAAGTGAATA 44681
QY 481 ProPheGlnLeuArgArg 486
Db 44680 CCGTTTCGGCCTATGCGG 44663
RESULT 2
AF458051 1834 bp DNA linear BCT 02-MAR-2002
LOCUS Pseudomonas syringae pv. maculicola type III effector HopToaPma
DEFINITION gene, complete cds.
ACCESSION AF458051
VERSION AF458051.1 GI:19071505
KEYWORDS
SOURCE Pseudomonas syringae pv. maculicola.
ORGANISM Pseudomonas syringae pv. maculicola
Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
Pseudomonas.
REFERENCE 1 (bases 1 to 1834)
AUTHORS Guttman,D.S., Vinatzer,B.A., Sarkar,S.F., Ranall,M.V., Kettler,G.
and Greenberg,J.T.
TITLE A functional screen for the type III (Hrp) secretome of the plant
JOURNAL pathogen Pseudomonas syringae
MEDLINE Science 295 (5560), 1722-1726 (2002)
PUBMED 21862332
11872842
REFERENCE 2 (bases 1 to 1834)
AUTHORS Vinatzer,B.A., Guttman,D.S., Sarkar,S.F., Ranall,M.V. and
Greenberg,J.T.
TITLE Direct Submission
JOURNAL Submitted (13-DEC-2001) MGCB, The University of Chicago, 1103 E
57th Street, Chicago, IL 60637, USA
FEATURES
source location/Qualifiers
1. 1834
/organism="Pseudomonas syringae pv. maculicola"
/strain="ES4326"
/db_xref="taxon:59511"
/note="pathovar: maculicola"
91. 116
/note="hrp box"
280. 1734
/note="similar to Pseudomonas syringae pv. tomato HopToaI
encoded by GenBank Accession Number AF232004"
/codon_start=1
/transl_table=11
/product="type III effector HopToaPma"
/protein_id="AAL84253.1"
/db_xref="GI:19071506"
/translation="MYINRISISSQSIGTESFHSQAQSVASSSVRLSEAEQTKLNDIT
HYLTDHVEFAHRLPVNDINVDQDVYHAHNEIGKIISARALRLDEGETALSTGEFA
KAEKEDMASAASGALRAIPEAAASVLOQMOPAIKGDMLPTPLKPLAIVSGALSGV
MDQVGTGMNRATGDLHYLSTAPEKLHDAMAASVKKRKGVMRQAADLGIAVQTYTAR
NALRTVLAPALASRPAYOGAVDISVASAGSLVANAGFSESMHTVOARDHLRGAFYLG
IKDKQPKADLSEETDWDLDAYRAIKSASYSGAALNAGRYAGLPIDVATDGLKAVRSIV
SATSILMONGIVMAGGFAGVGLQEMATKNITHPATKAAYSQJLNLGASAPVFSAMTTA
AAATDPAVKKAESFLQDPTVKTIVSNGTSLADKTVKLAAGIDASAERIATTGASLRD
TLRRRTVREPDIEEGVAAGTGSPPAVPFAMRP"

BASE COUNT 396 a 536 c 540 g 362 t
ORIGIN

Alignment Scores:
Pred. No.: 8.63e-91 Length: 1834
Score: 1591.50 Matches: 330
Percent Similarity: 78.07% Conservative: 51
Best Local Similarity: 67.62% Mismatches: 100
Query Match: 66.12% Indels: 7
DB: 1 Gaps: 3
US-09-825-414-66 (1-487) x AF458051 (1-1834)
QY 1 MethIleAsnGlnSerAlaGlnGlnProGlyValAlaMetGluSerPheArgThr 20
Db 280 ATGTACATCAACCGCTCCATCATGTTCTCAGTCCAGTATGTCACGAGAGCTTCATTCT 339
QY 21 AlaSerAspAlaSerLeuAlaSerSerSerValArgSerValSerThrThrSerCysArg 40
Db 340 GCGCAG-----TCTGTGCTCCAGTTCGGTCCAGGTGCTCACTGAGGCCGAGCAGACC 393
QY 41 AspleuGlnAlaIleThrAspTyrIleuLysHisValPheAlaAlaHisArgPheSer 60
Db 394 AAACGTGAATGACATCAACCCATTACCTGACAGATCATGTATTCCGCCACACAGGTTACCG 453
QY 61 ValIleGlySerProAspGluArgAspAlaAlaLeuAlaHisAsnGluGlnIleAspAla 80
Db 454 GTAAACGATATCAACGTTGATCATGAGATGTCGTTTCATGCCACACAGTCGAGATCGGCCAA 513
QY 81 LeuValGluThrArgAlaAsnArgLeuTyrSerGluGlyGluThrProAlaThrIleAla 100
Db 514 ATAAATCAGCGCCCGCTGCCCTACGCTTGCTTGATGAGGGCGGAACCGCTCTCAGCATTTGCC 573
QY 101 GluThrPheAlaLysAlaGluLysPheAspArgLeuAlaThrThrAlaSerSerAlaPhe 120
Db 574 GAAACCTTGCCCAAGGACAGAGAAGTTCAGACCGCATGGCAGATGCGGGCTCGGGCGCTTAA 633
QY 121 GluAsnThrProPheAlaAlaAlaSerValleuGlnTyrMetGlnProAlaIleAsnLys 140
Db 634 AGGCAACACCGCTTCGCGCGCGCGCTGCTGCAATACATGACGACCTGCCATCAACAAG 693
QY 141 GlyAspTyrLeuAlaThrProLeuLysProLeuThrProLeuIleSerGlyAlaLeuSer 160
Db 694 GCGACGTGTTACCGACGCGCGCTCAAGCCCTTGCGCCGCTGTCGCGCGCTATGCG 753
QY 161 GlyAlaMetAspGlnValGlyThrLysMetMetAspArgAlaArgGlyAspLeuHisTyr 180
Db 754 GCGGTCAATGACAGAGTTCGCGACCGCGCTCATGAACCGCGACCGGACCTGCATTAC 813
QY 181 LeuSerThrSerProAspLysLeuHisAspAlaMetAlaValSerValLysArgHisSer 200
Db 814 CTGAGCACATGCGCCGGAATAAATCTGCACATGCGATGGCTGCCCTGCTGTAATAACGCCATTAG 873
QY 201 ProAlaLeuGlyArgGlnValValAspMetGlyIleAlaValGlnThrPheSerAlaLeu 220
Db 874 CCGGTTGTCATGCGTCAAGGCTGCGGATCTGGGATCGCGCTTCAGACCTACACGGCGCGA 933
QY 221 AsnValValArgThrValLeuAlaProAlaLeuAlaSerArgProSerValGlnGlyAla 240
Db 934 AATGCGCTGCGTACCGTGTGCTGCTGCTGCACTGGCCCTCAGACCTGCAGTGCAGGTGCG 993
QY 241 ValAspPheGlyValSerThrAlaGlyGlyLeuValAlaAsnAlaGlyPheGlyAspArg 260
Db 994 GTAGACATCAACGCTGTCGCGCGCGGAGTACCTGTTGCCAACGAGGTTTATAGCAGCGC 1053
QY 261 MetLeuSerValGlnSerArgAspGlnLeuArgGlyGlyAlaPheValLeuGlyMetLys 280
Db 1054 ATGCATACGGGTGACGCGCGCTGATCACCTGCGTGGCGGTGCTTTGTACTCGGCATCAAG 1113
QY 281 AspLysGluProLysAlaAlaLeuSerGluGluThrAspTyrPheAspAlaTyrLysAla 300
Db 1114 GACAACAACCCCAAGGACAGACTTGAGTGAAGAAACCGACTGGCTCGATGCCACGGGCG 1173
QY 301 IleLysSerAlaSerTyrSerGlyAlaAlaLeuAsnAlaGlyLysArgMetAlaGlyLeu 320

QY	Db	Accession	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
QY	1174	ATCAAAATCCGCCAGTACTCCGGGTGCAGACGGCTCAACCGCTGCCAAGCGGTGTGCGCGGCTCTG																
QY	321	ProLeuAspValAlaThrAspGlyLeuLysAlaValArgSerLeuValSerAlaThrSer																
Db	1234	CCGCTGGACGCTGCCACCGATGGTTGAAAGCGGTGAGAGTCTGGTGTGGGCCACGAC																
QY	341	LeuThrLysAsnGlyLeuAlaLeuAlaGlyGlyTyrAlaGlyValSerLysLeuInds																
Db	1294	CTGATGCAAAACGGTCTGGTATATGGCGGGTGTGTTTCCCGGTGTGGTAAGTTGCAGAA																
QY	361	MetaLThrLysAsnIleThrAspSerAlaThrLysAlaAlaValSerGlnLeuSerAsn																
Db	1354	ATGGCGACGAAAAATATCACCCACCCGCGCAACCAAGGACCGCTCAGCCACTGACCAAT																
QY	381	LeuValGlySerValGlyValPheAlaGlyTyrThrThrAlaGlyLeuAlaThrAspPro																
Db	1414	CTTGACAGTTTCAGCCCCGGTTTTCAGCCTTGACACGCGCAAGCGGTGGCGACGACCCA																
QY	401	AlaValLysLysAlaGlySerPheIleGlnAspLysValLysSerThrAlaSerSerThr																
Db	1474	GCGGTGAAAAAGGCTGAGTCTGTTCTTGACAGGACACGCTGAAAAACAACGATATCCAACGG																
QY	421	ThrSerTyrValAlaAspGlnThrValLysLeuAlaLysThrValLysAspMetSerGly																
Db	1534	ACCAGCGCGCTGGCCGACAAAAACCGTCAAGCTGGCCAGGCGGCATTGACGCGACGCC																
QY	441	GLuAlaIleSerSerThrGlyAlaSerLeuArgSerThrValAsnLeuArgHisArg																
Db	1594	GAAAGGATGGCCACTACCGGGGGCTCCTTGCCTGACACC-----TTGCCCGCTGC																
QY	461	SerAlaProGluAlaAspIleGluGlyGlyIleSerAlaPheSerArgSer-----																
Db	1645	ACGGTGGCTGAGCCTGATATCGAGAGGGCGGTGTGCCGGCGGCACGGGTTGCCGGTT																
QY	479	GluThrProPheGlnLeuArgArg																
Db	1705	GCAGTGCCCTTTCGAGCCAATGCGG																
RESULT 3																		
AL646083																		
LOCUS																		
DEFINITION																		
ACCESSION																		
VERSION																		
KEYWORDS																		
SOURCE																		
ORGANISM																		
REFERENCE																		
AUTHORS																		
TITLE																		
JOURNAL																		

http://sequence.toulouse.inra.fr/R.solanacearum.html.

location/Qualifiers

1..208050

/organism="Ralstonia solanacearum"

/strain="GM1000"

/db_xref="taxon:305"

/plasmid="megaplasmid"

135..947

/gene="RSp1115"

/note="RS05505"

135..947

/gene="RSp1115"

/function="miscellaneous; not classified regulator"

/note="Product confidence : putative

Gene name confidence : hypothetical

predicted by Codon_usage

predicted by Homology

predicted by Framed"

/codon_start=1

/evidence=not_experimental

/transl_table=11

/product="PUTATIVE TRANSCRIPTION REGULATOR PROTEIN"

/protein_id="CAD18266.1"

/db_xref="GI:17431589"

/translation="MADSVQYQOVADVPGVLISAGRFSSQSFSEFHHFHDVHGLVTAG

VQRQRFKGTVLPGCISLMPGEIHDGVAEDGAYTKTFRLSQALLANLAQDISG

TDHEPELAATLEDPAIGHLRLDAMQHSASTLQVQSEMLALAEFLKOSRALIP

EHTKGALSPAQWQVRDYCFAHAGERITLDDIASLCAIGRFQFLKQRTVGMTPHAW

LVRRLRLEQACSLSSRSKAIAEVAQEVGFYDQSHFNRAFRQAFGVAPSHYQS"

1162..2259

/gene="RSp1116"

/note="RS05504"

1162..2259

/gene="RSp1116"

/EC_number="4.1.2.5"

/function="small molecule metabolism; degradation; amino

acids"

/note="Product confidence : putative

Gene name confidence : hypothetical

predicted by Codon_usage

predicted by Homology

predicted by Framed"

/codon_start=1

/evidence=not_experimental

/transl_table=11

/product="PUTATIVE LOW SPECIFICITY L-THREONINE ALDOLASE

PROTEIN"

/protein_id="CAD18267.1"

/db_xref="GI:17431590"

/translation="MNRDPMPTTRTPALGFTSDNIEGASPEVLDAILANAGQAAPYG

ADDCTARVEQTLSELEFEKVKFLVPTGTANALCIGAMPWGNVCHPSSHINDE

CGAPEFYTGAKLVAVDGPSSAKIDYEKLRATRVKGVGDVHSTQPAVSLTQATEVGS

YTLNEIEAIGQLCKDARIKLQMDGSRFANALVSLGCSPAAMTWKAGDALSFQATKNG

VVAEAIVLFDLSLAEMGYRRKRAGHLFSKMRFLSAQIEAYLANDLMLRNARQANDM

ALRLARGLGEGMHGASVIGATEANIVFCRLPSYLIENLLQAGFRFYHNREPNVRFVT

SFATTADVDVNDLRLRHKMAAAGPSADTHATA"

2257..5019

/note="ACUR or RS05503

predicted by Framed"

/evidence=not_experimental

complement(2462..3139)

/gene="RSp1117"

/note="RS05502"

complement(2462..3139)

/gene="RSp1117"

/function="miscellaneous; unknown"

/note="Product confidence : putative

Gene name confidence : hypothetical

predicted by Codon_usage

predicted by Framed"

/codon_start=1

/evidence=not_experimental

/transl_table=11

/product="PUTATIVE TRANSMEMBRANE PROTEIN"
 /protein_id="CAD18268.1"
 /db_xref="GI:17431591"
 /translation="MNDASLHETIISCLEHORPTVREIAERENRDEAQVRQALRALADNHGVLLHPHSDEWVIAHPESAPTTCVKSNGNRKMGNCAMCAGLVHLAGSATIETRIGAIDDHVILIRIENGRLDITDYVHFPIPMKHAMDNVITCSYQLLRFDEAQVDEWCASRGIPKGDVPRPIEQWNEFAVEWYARHADADWTKWSVREALTELFARHHLSGPVWTIAAEARF"
 complement(3188. .3595)
 /gene="RSp1118"
 /note="RS05501"
 complement(3188. .3595)
 /gene="RSp1118"
 /function="miscellaneous; hypothetical/global homology"
 /note="Product confidence : hypothetical
 Gene name confidence : hypothetical
 predicted by Codon_usage
 predicted by Homology
 predicted by Framed"
 /codon_start=1
 /evidence=not_experimental
 /transl_table=11
 /product="CONSERVED HYPOTHETICAL PROTEIN"
 /protein_id="CAD18269.1"
 /db_xref="GI:17431592"
 /translation="MEDTKIAIVRRDIDLATWOKLNTAFLMSGITGASPEIMGEAYRDAAADNTYHSLCQPVIVSGDETLLKAVHRRALERKVPGLYIEDMFSTGHDANRAVFAEHGPD SARVVGIALRREKKIVDKITGARHMP"
 complement(3673. .4161)
 /gene="RSp1119"
 /note="RS05500"
 complement(3673. .4161)
 /gene="RSp1119"
 /function="miscellaneous; hypothetical/global homology"
 /note="Product confidence : hypothetical
 Gene name confidence : hypothetical
 predicted by Codon_usage
 predicted by Homology
 predicted by Framed"
 /codon_start=1
 /evidence=not_experimental
 /transl_table=11
 /product="CONSERVED HYPOTHETICAL PROTEIN"
 /protein_id="CAD18270.1"
 /db_xref="GI:17431593"
 /translation="MGWSEVGETVCPRIARTLAVVGDRWTVLIRELFLGVKRFEEFQAQTGMSSHLLSTRLKRLDADGIVVRHLYSDRPRHEYRLTSGIDFVPLLSLKSWEKMGGFKAKTAPALTTIHRQGHETGLKLVCPACDEPFPGKDAYITLIGASFKAERQARDGA"
 4228. .4725
 /gene="RSp1120"
 /note="RS05499"
 4228. .4725
 /gene="RSp1120"
 /function="miscellaneous; hypothetical/global homology"
 /note="Product confidence : hypothetical
 Gene name confidence : hypothetical
 predicted by Codon_usage
 predicted by Homology
 predicted by Framed"
 /codon_start=1
 /evidence=not_experimental
 /transl_table=11
 /product="CONSERVED HYPOTHETICAL PROTEIN"
 /protein_id="CAD18271.1"
 /db_xref="GI:17431594"
 /translation="MPAITRHRPRGAPRYLDEFKVGDLTETSGFAVTRDMILSFAEQYDQPMHLDETARDVFGELVSGQFLAVTMRLLDARLLGGTPPIVGAELRDIRFHA PMRPGDVLARAEEVLAIRPSKSRPDRGFM DARVTTTNSGVTILVTQRWSLVVPRPDEFQKVAF"
 complement(5017. .5610)
 /gene="RSp1121"
 /note="RS05498"

CDS		complement(5017..5610) /gene="RSP1121" /function="miscellaneous; not classified regulator" /note="Product confidence : putative Gene name confidence : hypothetical predicted by Codon_usage predicted by Homology predicted by Framed" /codon_start=1 /evidence=not_experimental /transl_table=1 /product="PUTATIVE TRANSCRIPTION REGULATOR PROTEIN" /protein_id="CAD18272.1" /db_xref="GI:17431595" /translation="MAEKITRKPRADAEENRRQLLDVAKTAFEKGVASLEDIA AGVGIGLYRHFPTRDALIDEIYRDEGSRLEAARQSEDPLEAVRWLLFVG ANKQIRADVINCMTDCSERLCTISGEVLIEITLAOLIERAKQSAGTGLAVEPDLIS AGVASFGELDWEAGARLRIVEVMVAGLRVGAATGGKR" 5822..6910 /gene="RSP1122" /note="RS05497"
gene		
Alignment Scores:		
Pred. No.:	2.43e-17	Length: 208050
Score:	460.00	Matches: 132
Percent Similarity:	48.35%	Conservative: 58
Best Local Similarity:	33.59%	Mismatches: 153
Query Match:	19.11%	Indels: 51
DB:	1	Gaps: 9
US-09-825-414-66 (1-487) x AI646083 (1-208050)		
QY 30	SerValArgSerValSerThrThrSerCysArgAspLeuGlAlalIethrAspTyrLeu	49
Db 203624	:::: GCCGTGGCATCGACATCCCCACCTGGATGCCGATA-GCCGACCAGCTGC CGCATATTTC	::: 203682
QY 50	LysHisHisValpheAlaIahisArgpheserValIleGlySerProaspGluArgasp	69
Db 203683	::: AGGCCGCACCTTGACGGT-----	----- 203700
QY 70	AlaAlaLeuAlaHisasnGlnIleAspAlaLeuValGluThrArgAlaasnArgLeu	89
Db 203701	:::::: :::: ::::: -----ACCAACGATGGCGCAATTGCAGATCCTGATCCAGCAGCGCGCAGACGGCTC	203751
QY 90	TyrSerGluGlyGluThrProAlaThrIleAlaGluThrPheAlaLysAlaGluYsphe	109
Db 203752	::: ::: CAACGATGGGGAAACGCCGCAACCGTCGCCCGCTACTGCGCAAGGGCGCCAACCGG	203811
QY 110	AspArgLeuAlaThrThrAlaSerSerAlaPheGluAsnThrProPheAlaAlaIasr	129
Db 203812	:::: GACCGGCTGGCCACGACCAAGGTAGGCTTCGTGCGCTGGTGGCTTCGCGCATGGCTCG	203871
QY 130	ValLeuGlnTyrMetGlnProAlaIleAsnLysGlyAspTrpLeuAlaThrProLeuLys	149
Db 203872	:::::: ::: CGCCTGTTCGACGTGAAGCAGCACCTG-----ACGGCCTTCGCC	203910
QY 150	ProLeuThrProLeuIleSer-----GlyAlaLeuSerGlyAlaMetaspGlnVal	166
Db 203911	:::: AAGACGACGGCCGAGTCGGCGCGCAGCGTGGTGGCGGGCACGGCGATGGCTTC	203970
QY 167	GlyThrLysMetMetaspArgAlaArgGlyAspLeuHisTyrLeuSerThrSerProasp	186
Db 203971	:::::: :: :::: ::::: GGGGGCACCCCTGCTGGGCAAGCCACAGCAACACAATGGCTGGCGGCTCGCCGAT	204030
QY 187	LysLeuHisaspAlaMetaIaValserValLysArgHisSerProAlaLeuGlyArgGln	206
Db 204031	:::: ::: CACCTCGAACCTGTCAATGGCGCAAGCGCATAGAAGCCGTCAGCCCATCTGGGGCGGCTG	204090
QY 207	ValValAspMetGlyIleAlaValGlnThrPheSerAlaLeuAsnValValArgThrVal	226
Db 204091	::::: :: :: :: GCTGCGGAGGTGAGCCTGGCATTCACAGACCTAATTCCTGCGCAACGTGATCCGACCGGC	204150
QY 227	LeuAlaPro-----AlaLeuAlaSerArgProSerValGlnglyAlaValasp	242

Db 204151 GTCCGCCGCTGGCAACCATGCCCTGGGGCGGCCAAGGGCAGCCCAAC-----GTGCAT 204204
QY 243 PheGlyValSerThrAlaGlyGlyLeuValAlaAsnAlaGlyPheGlyAspArgMetLeu 262
Db 204205 TCCTGCATCCGGCGGTGGGGCGCCCTGTGCCGCTGCAGCATGCCATGCAG 204264
QY 263 SerValGlnSerArgAspGlnLeuArgGlyGlyAlaPheValLeuGlyMetLysAspLys 282
Db 204265 CACATGATAGAGACCACCACCGCAGCGGGCGGAATACCTGCTGGGC----- 204312
QY 283 GluProLysAlaAlaLeuSerGluGluThrAspTrpLeuAspAlaTyrLysAlaIleLys 302
Db 204313 -----CGCACCGACTGGGAAGACCAAGTCAAGCAGCTCAAG 204348
QY 303 SerAlaSerTyrSerGlyAlaAlaLeuAsnAlaGlyLysArgMetAlaGlyLeuProLeu 322
Db 204349 CAATCGACGTGACTGACCCCGCTGGTGGAGCGGGCAAGCGCACGCAAGCTGCTCGTC 204408
QY 323 ---AspValAlaThrAspGlyLeuLysAlaValArgSerLeuValSerAlaThrSerLeu 341
Db 204409 GCCGATTGGCGCAGCCGAGACCGCTGGCCGCCACGCGCAGCCTGTTCAACCGCCACCAACCTC 204468
QY 342 ThrLysAsnGlyLeuAlaLeuAlaGlyGlyTyrAlaGlyValSerLysLeuGlnLysMet 361
Db 204469 ATCAAGACATGGCGCGCAGCTGGCGGCTTGGCCGCTGTGTAACCGCACAGACCGCT 204528
QY 362 AlaThrLysAsnIleThrAspSerAla---ThrLysAlaAlaValSerGlnLeuSerAsn 380
Db 204529 GCCGGCAGAGCCGCCACGAGCGCGCTATACCGAGCGCGCGCCGCCCTCAGCGCGC 204588
QY 381 LeuValGlySerVal-----GlyValPheAlaGlyTrpThrThrAlaGlyLeuAla 397
Db 204589 GCCGTCAACACGGTACTGTGCGCACCGGTCTACGCTGCGTGACACAACCGCGATGTCTATG 204648
QY 398 ThrAspProAlaValLysLysAlaGlnSerPheIleGln 410
Db 204649 GCCGGCCCGCCATCGACGCCCGCGCGGCACATCCAG 204687

RESULT 4

AY094894/c 1558 bp mRNA linear INV 15-APR-2002
LOCUS Drosophila melanogaster RE67967 full insert cDNA.
DEFINITION AY094894
ACCESSION AY094894.1 GI:20151774
VERSION FLI_CDNA.
KEYWORDS Drosophila melanogaster.
SOURCE Drosophila melanogaster.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.

REFERENCE AUTHORS

1 (bases 1 to 1558)
Stapleton,M., Brokstein,P., Hong,L., Agbayani,A., Carlson,J.,
Champe,M., Chavez,C., Dorsett,V., Dresnek,D., Farfan,D., Frise,E.,
George,R., Gonzalez,M., Guarin,H., Krommiller,B., Li,P., Liao,G.,
Miranda,A., Mungall,C.J., Nunoo,J., Pacleb,J., Paragas,V., Park,S.,
Patel,S., Phouanenvong,S., Wan,K., Yu,C., Lewis,S.E., Rubin,G.M.
and Celniker,S.

TITLE JOURNAL

Submitted (03-APR-2002) Berkeley Drosophila Genome Project,
Lawrence Berkeley National Laboratory, One Cyclotron Road,
Berkeley, CA 94720, USA
Sequence submitted by:
Berkeley Drosophila Genome Project
Lawrence Berkeley National Laboratory
Berkeley, CA 94720

COMMENT

This clone was sequenced as part of a high-throughput process to
sequence clones from Drosophila Gene Collection 1 (Rubin et al.,
Science 2000). The sequence has been subjected to integrity checks
for sequence accuracy, presence of a polyA tail and contiguity
within 100 kb in the genome. Thus we believe the sequence to
reflect accurately this particular cDNA clone. However, there are
artifacts associated with the generation of cDNA clones that may

have not been detected in our initial analyses such as internal
priming, priming from contaminating genomic DNA, retained introns
due to reverse transcription of unspliced precursor RNAs, and
reverse transcriptase errors that result in single base changes.
For further information about this sequence, including its location
and relationship to other sequences, please visit our web site
(<http://fruitfly.berkeley.edu>) or send email to
cdna@fruitfly.berkeley.edu.

FEATURES source

1..1558
Location/Qualifiers
/organism="Drosophila melanogaster"
/strain="y; cn bw sp"
/db_xref="taxon:7227"
/map="64C5-64C5"
1..1558
/gene="CG13705"
/note="alignment with genomic scaffold AE003482"
/db_xref="FLYBASE:FBgn0035582"
59..1357
/gene="CG13705"
/note="longest ORF"
/codon_start=1
/product="R67967p"
/protein_id="AAM11247.1"
/db_xref="GI:20151775"
/db_xref="FBgn0035582"
/translation="MKFVIAFAVIAAASIASIATDYLPPVDNNLESASELVEVPAEQ
GVLSDGYRYKTVRRRLKLRHRDYNELPSGEYLPVEESASEAVAVETPLADGTRYK
TVRRVIRRRDVSSELSPEYLPVEESASEAVAVETPLADGTRYKTVRRVIRRRDVS
ELSPYLPVEESASEAVAVETPLADGTRYKTVRRVIRRRDVSSELSPEYLPVEES
ASEAVAVETPLADGTRYKTVRRVIRRRDVSSELSPEYLPVEESASEAVAVETPLA
ADGTRYKTVRRVIRRRDVSSELSPEYLPVEESASEAVAVETPLADGTRYKTVRRVI
RRRDVSELSPYLPVEESASEAVAVETPLADGTRYKTVRRVIRRRDVSSELSPE
YLPVEETASALIDVPAEQTVLASDGYRYKTVRRRLKLRH"

CDS

BASE COUNT 298 a 483 c 404 g 373 t
ORIGIN

Alignment Scores:
Pred. No.: 0.185 Length: 1558
Score: 160.50 Matches: 97
Percent Similarity: 38.52% Conservative: 59
Best Local Similarity: 23.95% Mismatches: 157
Query Match: 6.67% Indels: 92
DB: 3 Gaps: 15

US-09-825-414-66 (1-487) x AY094894 (1-1558)

QY 141 GlyAspTrpLeuAlaThrProLeuLysProLeuThrProLeuIleSerGlyAlaLeuSer 160
Db 1363 GCGCGCTTAGTGGCGACGAGCTTCAGACGAGCAGCGGTCTTGTAGCGGTAGCCATCGCT 1304
QY 161 GlyAlaMetAspGlnValGlyThrLysMetMetAspArgAlaArgGlyAspLeuHis--- 179
Db 1303 GCGCAGACGGTCTGCTCGCGGGCACATCATGATGCGACAGCGCGTCTCTCCACGGG 1244
QY 180 -----TyrLeuSerThrSerProAspLysLeu 188
Db 1243 GGGCAGTACTCGCCGGAAGGCAGCTCAGACATCACGGCGCGGACGAGTACACAGCAGC 1184
QY 189 HisAspAlaMetAlaValSerVal-----LysArgHisSerProAlaLeuGly 204
Db 1183 CACGGTCTTGTAGCGGTATCATCATCAGCCAGGGAGAGTCTCCACAGCAAGCCTCGGA 1124
QY 205 ArgGlnValValAspMetGlyIleAlaValGln-----ThrPheSer 218
Db 1123 GCGGACTCCTCAACAGGGGCGAGTATTCAGGGGACAGCTCGAGACATCACGACGACG 1064
QY 219 AlaLeuAsnValValArgThrValLeuAlaProAlaLeuAlaSerArgProSerValGln 238
Db 1063 ACGGATCACACGACGACGCGGTCTTGTAGCGGTATCCATCATCAGCCACAGAGAGTCTGCAC 1004
QY 239 GlyAlaValAspPheGlyValSerThrThrAlaGlyGlyLeuValAlaAsnAlaGlyPheGly 258
111

Db 1003 GGC-----AACGGCTTCGGA 989
QY 259 AspArgMetLeuSerValGlnSerArgAspGlnLeuArgGlyAlaIaphValLeuGly 278
Db 988 GCGCGATTCTC-----AACAGGGGGCGATTTCGGCGGACAGCTCGTTTAC--- 941
QY 279 MetLysAspLysGlnProLysAlaAlaLeuSerGlnGluThrAspTrpLeu----- 295
Db 940 -----ATCAGCAGCGCGCAGCATCACAGCAGCAGCAGCTTGTAGCGGTA 896
QY 296 -----AspAlaTrpLysAlaIleLysSerAlaSerTyrSerGlyAlaAlaLeuAsnAla 313
Db 895 TCCATCAGCAGCAGAGGGGTCTCAACAGCAGCAGCAGCAGCTCGGAGGAGCAGCTCCTCAACGGG 836
QY 314 GLYLSArgMetAlaGlyLeuProLeu---AspValAlaIleThrAspGlyLeuLysAlaVal 332
Db 835 GGGCAGGTACTCGCCGGAAGGAAGCTCATGTGACATCACGACGACGCGATCACACGGCG 776
QY 333 ArgSerLeuValSerAlaIleThrSerLeuThrLysAsnGlyLeu-----AlaLeuAla 349
Db 775 GACGGTCTTGTAAACGTTAACCATCATCATCAGCAAGAGGGGTGTCACAGCAACAGCAGCTCAGA 716
QY 350 GlyGlyTyrAlaGlyAlaSerLysLeuGlnLysMetAlaThrLysAsnIleThrAspSer 369
Db 715 GCGCGACTCTCTCAACAGGGGGCAGGTATTTCAGGGGAGCAGCTCGAAACATCACGACGACG 656
QY 370 AlaThrLysAlaAlaValSerGlnLeuSerAsnLeuValGlySerValGlyValPheAla 389
Db 655 ACGGATCACACGCGGAGAGTCTTGTAGCGGTAGCCATCATCACGAGGAGAGTCTCGAC 596
QY 390 GlyTrpThrThr---AlaGlyLeuAlaThrAspProAlaValLysLysAlaGlySerPhe 408
Db 595 GGCACACAGCCTCGAGGCGGACTCCTCAACAGGGGGCAGGTACTCAGGGAGAGTTTC--- 539
QY 409 IleGlnAspLysValLysSerThrAlaSerSerThrThrSerTyrValAlaAspGlnThr 428
Db 538 -----GGAGACATCACGACGACGACGATCACAGCGCGGACAGTCTT 497
QY 429 ValLysLeuAla-LysThrValLysAspMetSer-----GlyGluAlaIleSe 444
Db 496 GTAGCGGTAGCCATCATCATCAGGAGAGTCTCGACAGCAACAGCCTCGGAGGCGGACTC 437
QY 444 rSerThrGlyAla-----SerLeuArgSerThrValAsnAsnLeuArgHisArg----- 460
Db 436 CTCACACAGGGGCGAGTACTCAGGGGAGAGTTTCGAGACATCACGACGACGACGATCAC 377
QY 461 -----SerAl 462
Db 376 ACGGCGGAGCGTCTGTAAAGGTAGCCATCATCACGCCAGAGAGTCTCGACGCGCAACGGC 317
QY 462 aProGluAlaAspIleGlnGluGlyIleSerAlaPheSerArgSerGluThrProPh 482
Db 316 TTCGGAGGCGGATTCCTCAACAGGGGGC-----AGTATTTCACCGGAGGAGAGTCTGTT 263
QY 482 eGlnLeuArgArg 486
Db 262 CACATCACGACGG 250

RESULT 5
AC108877/c 163709 bp DNA linear INV 22-JUN-2002
LOCUS Drosophila melanogaster 3L BAC RP98-17L24 (Roswell Park Cancer
DEFINITION Institute Drosophila BAC library) complete sequence.

ACCESSION AC108877
VERSION AC108877.2 GI:21538999
KEYWORDS HTG.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster

Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 163709)
AUTHORS Muzny,D., Scherer,S., Adams,M.D., Holt,R.A., Evans,C.A.,

TITLE
JOURNAL
REFERENCE
AUTHORS

Gocayne,J.D., Tabor,P., Williamson,A., Homs1,F.H.,
Dugan-Rocha,S.D., Sodergren,E.S., Hodgson,A.H., Chen,R.C.,
Ayele,M., Scott,G.S., Worley,K.W., Amamatides,P.G., Brandon,R.C.,
Rogers,Y., An,H., Baldwin,D., Beeson,K.Y., Brown,M., Buhay,C.,
Busam,D.A., Center,A., Chen,G., Chen,Z., Clerc-Blankenburg,K.,
Davenport,L.B., Dietz,S.M., Ding,Y., Dodson,K., Doup,L.E.,
Draper,H., Emery-Cohen,A., Ferrleria,S., Garg,N.D.S., Houck,J.,
Hostin,D., Howland,T.J., Hume,J., Ibegwan,C., Jalali,M., Kovar,C.,
Liu,W., Mattei,B., McIntosh,T.C., Morgan,M., Moy,M., Murphy,B.,
Nelson,K.A., Ndassa,Y., Nguyen,N., Perez,L., Pittman,G.S., Puri,V.,
Schaefer,F., Shen,H., Strong,R., Tector,C., Wang,Q., Williams,S.M.,
Xiang,J., Zaveri,J.S., Zhou,J., Zorrilla,S., Smith,H.O.,
Wheeler,D., Weinstein,G., Gibbs,R. and Venter,J.C.

Direct Submission
Unpublished

2 (bases 1 to 163709)
Worley,K.C., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Banks,T., Barbara,J.,
Benton,J., Bimage,K., Blankenburg,K., Bonnin,D., Bouck,J.,
Bowle,S., Brlewa,M., Brown,E., Brown,M., Bryant,N.P., Buhay,C.,
Burgh,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carron,T.F.,
Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R.,
Chen,Z., Chowdhry,I., Christopoulos,C., Cleveland,C.D., Cox,C.,
Coyle,M.D., Dathorne,S.R., David,R., Davila,M.L., Davis,C.,
Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O.,
Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H.,
Dugan-Rocha,S., Durbin,K.J., Earnhart,C., Edgar,D., Edwards,C.C.,
Elhaj,C., Escotto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J.,
Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T.,
Garza,N., Gill,R., Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S.,
Hamilton,K., Harris,C., Harris,K., Hart,M., Haylak,P., Hawes,A.,
Hernandez,J., Hernandez,O., Hodgson,A., Hognes,M., Holloway,C.,
Hollins,B., Homs1,F., Howard,S., Huber,J., Hultky,S., Hume,J.,
Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S.,
Joudah,S., Karlsson,E., Kelly,S., Khan,U., Kling,L., Korvah,J.,
Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,
Lewis,L., Li,J., Li,Z., Licharge,O., Lieu,C., Liu,J., Liu,W.,
Louisege,H., Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,
Ma,J., Maheshwari,M., Mapua,P., Martin,R., Martindale,A.,
Martinez,E., Massey,E., Mawhney,E., McLeod,M.P., Meador,M.,
Mei,G., Metzker,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,
Morgan,M., Morris,S., Moser,M., Neal,D., Newton,J., Newton,N.,
Nguyen,A., Nguyen,N., Nguyen,N., Nickerson,E., Nwokenkwo,S.,
Oguh,M., Okwuonu,G., Oragunye,N., Oviedo,R., Pace,A., Payton,B.,
Peery,J., Perez,L., Peters,L., Pickens,R., Primus,E., Pu,L.L.,
Quiles,M., Ren,Y., Rives,M., Rojas,A., Rojubokan,I., Rolfe,M.,
Ruiz,S., Savery,G., Scherer,S., Scott,G., Shen,H., Shooshitari,N.,
Sisson,I., Sodergren,E., Sonaike,T., Sparks,A., Stanley,H.,
Stone,H., Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K.,
Tang,H., Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N.,
Thomas,S., Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R.,
Wall,R., Wang,S., Ward-Moore,S., Warren,R., Washington,C.,
Watlington,S., Williams,G., Williamson,A., Wleczyk,R., Wooden,S.,
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
Weinstock,G. and Gibbs,R.

TITLE
JOURNAL
REFERENCE
AUTHORS

Unpublished
Direct Submission
3 (bases 1 to 163709)
BCM-HGSC.

Direct Submission
Submitted (01-FEB-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

REFERENCE
AUTHORS
TITLE
JOURNAL

Direct Submission
Submitted (22-JUN-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA
On Jun 22, 2002 this sequence version replaced gi:18463997.
INFORMATION: <http://www.hgsc.bcm.tmc.edu/> or email
gc-help@bcm.tmc.edu

segment 3/11.
AL646078 AL646053
AL646078.1 GI:17430778
.
Ralstonia solanacearum.
Ralstonia solanacearum
Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
Ralstonia.
1 (bases 1 to 203050)
Salanoubat,M., Genin,S., Artiguenave,F., Gouzy,J., Mangenot,S.,
Arlat,M., Billault,A., Brotlier,P., Camus,J.C., Catolico,L.,
Chandler,M., Choisine,N., Claudel-Renard,C., Cunnac,S., Demange,N.,
Gaspin,C., Lavie,M., Moisan,A., Robert,C., Saurin,W., Schlex,T.,
Sigquier,P., Thebault,P., Whalen,M., Wincker,P., Levy,M.,
Weissenbach,J. and Boucher,C.A.
Genome sequence of the plant pathogen Ralstonia solanacearum
Unpublished
2 (bases 1 to 203050)
Boucher,C.A.
Direct Submission
Submitted (05-DEC-2001) Genoscope and CNRS UMR-8030, 2 rue Gaston
Cremieux, CP5706, 91057 Evry Cedex, France, Laboratoire de Biologie
Moléculaire des Interactions Plantes-Microorganismes INRA-CNRS,
BP27, 31326 Castanet-Tolosan Cedex, France, Fondation Jean
Dausset-CEPH, 27 rue Juliette Dodu, 75010 Paris, France, ILMG CNRS
118 Route de Narbonne, F 31062 Toulouse Cedex 4, Genoscope and INRA
URGV, 2 rue Gaston Cremieux, CP5706, 91057 Evry Cedex, France,
Laboratoire de Biometrie et Intelligence Artificielle INRA, BP27,
F31326 Castanet-Tolosan Cedex, Laboratoire de Genetique Cellulaire
INRA, BP27, F31326 Castanet-Tolosan Cedex
Christian.Boucher@toulouse.inra.fr
http://sequence.toulouse.inra.fr/R.solanacearum.html.
Location/Qualifiers
1. .203050
/organism="Ralstonia solanacearum"
/strain="GMI1000"
/db_xref="taxon:305"
/plasmid="megaplasmid"
complement(147. .1499)
/gene="RSp0310"
/note="RS05460"
complement(147. .1499)
/gene="RSp0310"
/function="cell processes; transport of small molecules"
/note="Product confidence : probable
Gene name confidence : hypothetical
predicted by Codon_usage
predicted by Homology
predicted by Framed"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="PROBABLE MACROLIDE-EFFLUX TRANSMEMBRANE PROTEIN"
/protein_id="CADI7461.1"
/db_xref="GI:17430779"
/translation="MNAPRKNTDRAEAALLRNASTTHEPANPCWIPDAAAAGLHPGR
WRARFWAIFGGQALSLIGSALITQFVLLMWITDTGVSALATAGLAALLPQAVLSPLG
GTFADRYSRRLMIVADGISALCMLLIALFLTGRIALMHAYVMAIRSAQAQFOAPA
AAASVTMLVPGGFLVRAGLSLSQSLTVVAAPLGALAGVMPIGWALGIDVATALL
GIAPLLCCRIPOACVSSGQKGLMPAFREGVDLVWKTGRLWLVLLGAVLTIIMPTF
TLVPLLVKTHFGGAAQVALLMEGLSIGMVAGGLVAAMAPRQVAMILCGFAMSCYA
IALTALAPGSLFGVAVAMWVISCITFEVFNAPLTALQTIYPNHLQGRVLSLNTTMG
LAAPVGLALFTPLGEVIGVRAIFAIALGLGAASLAGFLSPSLMRDRPSRSEARTD
"
gene
complement(1531. .2961)
/gene="oprM"
/note="RSp0311; RS05459"
complement(1531. .2961)
/gene="oprM"
/function="cell processes; transport of small molecules"
/note="Product confidence : probable
Gene name confidence : putative
predicted by Codon_usage

predicted by Homology
predicted by Framed"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="PROBABLE OUTER MEMBRANE CHANEL LIPOPROTEIN"
/protein_id="CADI7462.1"
/db_xref="GI:17430780"
/translation="MRRLPLVLIALSACSLPALIKPALVPATFPAESGPETQAAH
ADLGWRTMFGDRRLQRLIGLTLANRDLRLALNVEVTRAQYSIQRAARLPGIDATGS
ASRQRTAANDGASPAALANSEQYGVNGLNAFEIDLFGVRSLSDAFAFRLASDOG
BRAAQISLVGAVADAYFAERLAQEQROLAEHTLADWRQSLDARLKTAKONSDLDVA
QAEQVANAAQADLEARTBALAQAGNSLQRLVGAGLPADLPAPLSLETOPVYTQPLAGL
PSLLIRPDIRQAEHNLAAANADIGAARAAPFPRLSLTASLGYPAMGSLFDGCHR
VWSFSPQITQPLFGGRLRAELRLAELRKSATAEYERTIOFAFREVADGLAGATYG
RQIEAQIRAVASAERVELSNLRYRAGVEGRLELIDSQROLYASRQALLDLQGAFSN
AVALYKALGGGLAETDYATADRVARK"
complement(2966. .6115)
/gene="mexD"
/note="RSp0312; RS05457"
complement(2966. .6115)
/gene="mexD"
/function="cell processes; transport of small molecules"
/note="Product confidence : probable
Gene name confidence : putative
predicted by Codon_usage
predicted by Homology
predicted by Framed"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="PROBABLE MULTIDRUG EFFLUX SYSTEM TRANSMEMBRANE
PROTEIN"
/protein_id="CADI7463.1"
/db_xref="GI:17430781"
/translation="MPQFIRRPVFAWVIALFTILLGAIAIPKLPRIARPSVAPPTVN
LSATYPGATPQTMNDGVISLIERLSSVKNLYPESSSDTSGSASIVATFRKGTDPAM
AOVDVQNKLIKAIERLPQVYRONGLTVESASSGFLMIIGLQSDDGFRDETAISDYMAR
NVVEELRRIEGVGRVQLFGAEQAMRIWDPAKLIAGLSMNDLSAISQNAOIAFGS
VGSPALPAQRVTVPITVQGLTTPEOFAIVLRANANGSKVMGDVARELGAQSYG
FINRENGKATTAAVQDLAPGANAVRTAEGVQARNAELSRTPAGMSYSIPNTAPFVK
ISIEKVIOTLIEAMVLVEFLMYLFLQNVRYTLIPAIAPIALIGFTIMLASGSINV
LTFEGVLAIGIIVDAIVVENVERLMAQEGLSPRDATIKAKEITGAIVGITVLVT
AVETIPMALASGSVGIYQOFTLSMAVSIVEAFIALTLTPALCATLLKPVDAHHEKR
GVFGWENRRFDRITQGYAARVGLVTRGRVMLVFGAIAVLAALAEFRQLPSSFLPED
QGYFMTTFQLPADATAERTLDVYKAFERHVASRASISQNSIVGFGSGSNAAMAF
TMLKDWARNGATAQONEMRLAQQAQAGATEGTWSDLPPAIDELGTSSGFTLRQDRA
NQGDAALKAETKLELAOSKRIVSGVPPDGLPAGTSIRLIDIDRQAEALGVAFQTSIS
ETLSTAIGSLYVNDPFGNAGRMQOVIIOADAPARQIDDVLLKIDVRNAGGVPLSEVV
RPVWSETPLQLVRFQGFPAARISGNAPAGASSGAMAEMERLAAQLPPGFATIAWTQOS
LOEROSAAQAPMLMALSMLEVLELIALYESMALPLSMLVPLGLGAVLAIVLLRGM
PNDVFEKVGMITVIGLSAKNAILLIVEFAKQLREGQGLVESAVTAARLRRLPILMTSL
AFGLGVPLIMIAMGASAEQHAIGTGVFGMVSATVLAIFEVVFVFVFMGLQERFEA
WRASRQAPPPVPVPHQPEG"
complement(6136. .7293)
/gene="mexC"
/note="RSp0313; RS05456"
complement(6136. .7293)
/gene="mexC"
/function="cell processes; protection responses;
drug/analog sensitivity"
/note="Product confidence : probable
Gene name confidence : putative
predicted by Codon_usage
predicted by Homology
predicted by Framed"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="PROBABLE MULTIDRUG EFFLUX SYSTEM PROTEIN"
/protein_id="CADI7464.1"
/db_xref="GI:17430782"
/translation="MKIERLRHGGVHVLLAAMLACGCPSEPSQAPPAEVAATTIVAP

AHLELTEDLPGRVAAVRAEIRPOVSGIVORLFEQGETVRAGQPLFOINPAPERADA
DTAAALRRAEALARARVQTRLOPLVEADAVSROAYDDAVSQRDQAADVEQARAT
LARROLDKFATVEAPIGRIDOLVTEGALVSSGDSNPMARIOIDQVYDVRRPAS
SLEALROTLATQKTGAGDGLPVAVLRSNGEPEYAKGHILFSGINVDAGTGDVLLRVLV
NNPORQLPGMEVBARVPHARYTDALTVPQAVYRVGVGPOVMTLDANSHARLKPVEL
GELADRRYRIRAGLSAGQKIVEGMERLSDGAVAHAHDKWSPEVPAGPAR"
7392. .8162
/gene="Rsp0314"
/note="RS05454"
7392. .8162
/gene="Rsp0314"
/function="miscellaneous; not classified regulator"
/note="Product confidence : probable
Gene name confidence : hypothetical
predicted by Codon_usage
predicted by Homology
predicted by Framed"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="PROBABLE TWO COMPONENT RESPONSE REGULATOR
TRANSCRIPTION REGULATOR PROTEIN"
/protein_id="CAD17465.1"
/db_xref="GI:17430783"
/translation="MELNQGAAPVMPPSGAQSAQALLILIAEDEPETADILAA
YLARSLRSVHADGREALELHLSLKPDVLVDYQMPHVDGKALSEIRHGDTPVIM
LTALDQDIDKLMGLRIGADDVVKFPNPAEVARTQAVLRSMGSTRDRQVLRVAP
EQIDLNDHEATVOGQRHALALTLEFKLLAQLARAPKRVFSRAELLACLPEDDTL
ERTVDSHISKRLKLEDLGVNGMPVGVGVGRVLSLE"
8159. .9241
/gene="Rsp0315"
/note="RS05453"
8159. .9241
/gene="Rsp0315"
/EC_number="2.7.-.-"
/function="miscellaneous; not classified regulator"
/note="Product confidence : probable
Gene name confidence : hypothetical
predicted by Codon_usage
predicted by Homology
predicted by Framed"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="PROBABLE TRANSMEMBRANE SENSOR KINASE
TRANSCRIPTION REGULATOR PROTEIN"
/protein_id="CAD17466.1"
/db_xref="GI:17430784"
/translation="MKLAGLSRQIALSMMAAFGVTLVLVLTSYAFYYLAMTYWSRDF
CQASWLP TGPEWVWLIAITLQGLAMAIIVAVNLAIRILVPLNSVADSLRRVARGDLGV
RAIAGDRSLGEALADDEFNALADQLQRTVEQVFNAAIAHELRTPTVTLRGLQGL
AEGVTPDEPQFRSLTLQVESLTRLIEDLRVLSLAESGHLDLQVRESDLVAEVKAVVA

Alignment Scores:
Pred. No.: 516 Length: 203050
Score: 150.50 Matches: 123
Percent Similarity: 33.10% Conservative: 69
Best Local Similarity: 21.21% Mismatches: 220
Query Match: 6.25% Indels: 168
DB: 1 Gaps: 21
US-09-825-414-66 (1-487) x AL646078 (1-203050)
QY 2 HisIleasnGlnSerAlaGlnGlnProGlyVal-AlaMetGlnSerPheArgThrAl 21
Db 51178 CACCTACCTGTGCGACAGACAGCGGCGCTGGGCTCGGCGCTGACACGCTTCTTCATGCG 51237
QY 21 aSerAspAlaSerLeuAla-----SerSerSerValArgSerValSerThrTh 37
Db 51238 CGCGCGAC---AGCCTGGCCTCCATTCCTCCAGCACAGTCCCGCCAGGCTCTTCCCTGAG 51294
QY 37 rSerCysArgAspLeuGlnAlaIleThrAspTyrLeuLysHisHisValPheAlaAlaHis 57

Db 51295 CGCGCGAGCACCCCTGCAGACC-----
QY 57 sArgPheSerValIleGlySerProAspGluArgAspAlaAlaLeuAlaHisAsnGluG1 77
Db 51317 -CGCTCAACCGCATCTCGGGG-----CAGATGGCGTCCCTGTGCAAC-----CA 51360
QY 77 nIleAspAlaLeuValGluThrArgAlaAsnArgLeuTyrSerGluGluGluThrProAl 97
Db 51361 GGTCACACCCAGGTGCGAGACCGCAGGTCAACAGCGTCAACAGCACCAACAGCATCGC 51420
QY 97 aThrIleAlaGluThrPheAlaLysAlaGluLysPheAspArgLeuAlaThrThrAlaSe 117
Db 51421 CGCGCTCAACGACAGATCGCCAGGCCGAA-----GCCGCGAG 51459
QY 117 rSerAlaPheGluAsnThrProPheAlaAlaAlaSerValLeuGlnTyrMetGlnProAl 137
Db 51460 CGGCGCG-----CAGCCCGC 51474
QY 137 aIleAsnLysGlyAspTyrPleuAlaThrProLeuLysProLeuThrProLeuIleSerG1 157
Db 51475 CAATGACCTGCGCGACACAGCGGACAGCTGTGTGACAGCGCTCAATCAGTCGATCAAGGC 51534
QY 157 yAlaLeuSerGlyAlaMetAsp-----GlnValGlyThrLysMetMetAs 172
Db 51535 CAGCGTGTGCGACAGACAGCGGCGCAGTACACATCTACGTGCGCAACAGGCCAGCGCG-- 51592
QY 172 pArgAlaArgGlyAspLeuHisTyrLeuSerThrSerProAspLysLeuHisAspAlaMe 192
Db 51593 -CTGCTGACAGGCGCAACAGAGCTACAGCTGACGAGCGGTGCGATTCGCAATACGATCCGAC 51651
QY 192 tAlaValSerValLysArgHisSerProAlaLeuGlyArgGlnValValAspMetGlyI1 212
Db 51652 GCAGCTGTGCTGGTGGCTACAGAGCCGCGCGCACGCGTACCATCGACAGACAGCAGCT 51711
QY 212 eAlaValGlnThrPheSerAlaLeuAsnValValArg--ThrValLeuAlaProAlaLe 231
Db 51712 CGCGCGCGCGCGCTCGCGCGCTGATGCAGTTCGCCAGAACAGCTGATCCGCGCGCA 51771
QY 231 uAlaSerArgProSerValGlnGlyAlaValAspPheGlyVal----- 245
Db 51772 GAACAGCCTGGCGCGCGCTGCGCGCGCGCTCGCGCGAGCTCATACGACAGACAAGA 51831
QY 246 -----SerThrAlaGlyGlyLeuVa 252
Db 51832 GGGCATGACCTGAACGGCAAGCTGGGGACCGATCTGTCTCTCCACCGCGCGCCAGCGT 51891
QY 252 lAla-----AsnAlaGlyPheGlyAspArgMetLeuSerValGlnSerArgAs 268
Db 51892 CGCGCGCAGCTCCAGCAACACCGCGCGCGCGCTGACGGCGGACCATCACCAATGCCAA 51951
QY 268 p-----GlnLeuArgGlyAlaPheValLeuGly-- 278
Db 51952 CGCGCGCGCGGCTACGACTACAGAGTCAGATTCGGGGCGCGCTATACGGTGTGCGCA 52011
QY 279 -MetLysAspLysGluProLysAlaAlaLeuSerGluGluThrAspTyrPleuAspAla-- 297
Db 52012 TTACCCGAGCGGCTCCGCGCGCGTGAAGGTGTGAGCTGGCCGACACCGTCGACGCGCGT 52071
QY 297 ----- 297
Db 52072 CACGCTCAACCTGACGGGCTGATGGCCAGCGGCGACTCTCTCTGCTGCGGCCCAACCGT 52131
QY 298 -----TyrLysAlaIleLysSerAlaSe 305
Db 52132 CAACGGCGGATCGACCATGACAGCGCTGACCTCCGATTACACCGGCGGTGCGCGCGCC-- 52189
QY 305 rTyrSerGlyAlaAlaLeuAsnAlaGlyLysArgMetAlaGlyLeuPro-----LeuAs 323
Db 52190 ---TCGCGCGGTGTGTCAACCAAGGCGACACACACAGCGGCGACGTCGGTGGCGGTC 52245
QY 323 pValAlaThrAspGlyLeuLysAlaValArgSerLeuValSerAlaThrSerLeuThrLy 343
Db 52246 CATCGCGGTGACAGACACCTATGCCGCTCGCGCGCTCGCGCTCGCGCTCAGCCTGACCTA 52305

RDVFTRRLRYTLKGOTTFFDEIKRKYERBEAFRTAVDRYITDFEKLADVAARTDRDRSVT
OSYLSDTGKVVYTMALAHAGRFN"
BASE COUNT 1772 a 2913 c 3262 g 2082 t
ORIGIN

Alignment Scores:
Pred. No.: 10.8 Length: 10029
Score: 149.50 Matches: 96
Percent Similarity: 34.70% Conservative: 65
Best Local Similarity: 20.69% Mismatches: 176
Query Match: 6.21% Indels: 127
DB: 1 Gaps: 16

US-09-825-414-66 (1-487) x AE008061 (1-10029)

QY 69 AspAlaAlaLeuAlaHisasnGlunIleAspAlaLeuValGluThrArgAlaAsnArg 88
|||||
Db 8049 GATGCCCTTCAACCAATTCGACGCGGAAGATCGATACCGTTCTGGCGGAACGACGACGCC 7990
QY 89 LeuTyrSerGluGluThrProAlaThrIleAlaGluThrPheAlaLysAlaGluLys 108
|||||
Db 7989 CTGTTC-----GGCGCGCTTTCGGCCAGCCAGACCGCT 7957
QY 109 PheAspArgLeuAlaThrThrAlaSerSerAlaPhe-----GluAsnThrPro 124
|||||
Db 7956 TTCGACGAAACACTCGCGACCGCTTCTCTCGCCATTACCGGCTCGGTTCCGCGACGCGC 7897
QY 125 PheAlaAlaAlaSerValLeuGlnTyrMetGlnProAlaIleAsnLys----- 140
|||||
Db 7896 GAACACCTTGCCCGCATGCTCGATGAACGCGCGCGCGCATCAACAGCGTCGTGCCGAC 7837
QY 141 ---GlyAspTrpLeuAlaThrProLeuLysProLeuThrProLeuIleSerGlyAlaLeu 159
|||||
Db 7836 GTCGAACGTGCGCGATGCTCGATGAACGCGCGCGCGCATCAACGCGCGCTGTC 7777
QY 160 SerGlyAlaMetAspGlnValGlyThrLysMetMetAspArgAlaArgGlyAspLeuHis 179
|||||
Db 7776 TCCGGCATCGAGGACCGCATCTCCGACACGCTGAAACCGCGCACGCGCTGCC----- 7726
QY 180 TyrLeuSerThrSerProAspLysLeuHisAspAlaMetAlaValSerValLysArgHis 199
|||||
Db 7725 -----CTGCATGATGTCGTTTCCGGCGGAGAGCGCGCATC 7690
QY 200 SerProAlaLeuGlyArgGlnValValAspMetGlyIleAlaValGlnThrPheSerAla 219
|||||
Db 7689 GCCGATACGCTTGAC-----GGCGCGACCGCGCTCTTTCACGCGC 7648
QY 220 LeuAsnValValArgThrValLeuAlaProAlaLeuAlaSerArg----- 234
|||||
Db 7647 ATTTCCGGTGTGAGGAACGTATTGCCGACACGATGGACAGCCGACGCTGTCACTCGAC 7588
QY 235 -----ProSerValGlnGlyAlaValAspPheGlyValSerThr 247
|||||
Db 7587 ATGACCTTCGCCAATGTTGAGGAACGGCTTCCGAGACGCTCGACAACGCGACCTCCGCT 7528
QY 248 AlaGlyGlyLeuValAlaAsn----- 254
|||||
Db 7527 CTGACAGGCAATTGTCGAAGCGCCGAAGAAAGATCCCGGGCGCTCGATAGCGGACG 7468
QY 255 AlaGlyPheGlyAspArgMetLeuSerValGlnSerArg-----AspGlnLeuArgGly 272
|||||
Db 7467 GCCACCTTCGTTGACGTGGTTCGCGGTGCTGAACCCGTATCGCCGAAACGCTGATGGC 7408
QY 273 -----GlyAlaPheValLeuGlyMetLysAspLysGluProLysAlaAla 287
|||||
Db 7407 CGCACCGCAGCGCTCAATGCCGTGTTTCCGTCGAGGAGCGC----- 7363
QY 288 LeuSerGluGluThrAspTrpLeuAspAlaTyrLysAlaIleLysSerAlaSerTyrSer 307
|||||
Db 7362 -----ATCGCGGATGCTCTCGACAGCCGCTACCATGCGGCTCGACATGACCTTCTCC 7312
QY 308 GlyAla----- 309
|||||

Db 7311 GCGCGCCGAGGAAAGATCGCCGAAGCGCTCGACACGCGCACCGCCCTTGCTGAGCTG 7252
QY 310 AlaLeuAsnAlaGlyLysArgMetAlaGlyLeuProLeuAspValAlaThrAspGlyLeu 329
|||||
Db 7251 GTGGCAAGCGCCGAGACCCGTATTGCCGCGC---GGCGTCGACAGCCGACGATTCGCTG 7195
QY 330 LysAlaValArgSerLeuValSerAlaThrSerLeuThrLysAsnGlyLeuAlaLeuAla 349
|||||
Db 7194 AAAACCGTC----- 7186
QY 350 GlyGlyTyrAlaGlyValSerLysLeuGlnLysMetAlaThrLysAsnIleThrAspSer 369
|||||
Db 7185 -----GTTTCCGGCGCGGGAAGAGCGGATCACCC---GATGTGCTGACAGC 7144
QY 370 AlaThrLysAlaAlaValSerGlnLeuSerAsnLeuValGlySerValGlyValPheAla 389
|||||
Db 7143 CGCACCATGCGCGTCGACATGAGCTTCTCCGCGCTCGAGAAAGATCACCAGATTTCTC 7084
QY 390 GlyTrpThrThrAlaGlyLeuAlaThrAspProAlaValLysLysAlaGluSerPheIle 409
|||||
Db 7083 GATGCCCGCACCGCC-----GGCGTGAAGAGCGCGCTGCGGCTGT 7042
QY 410 GlnAspLysValLysSerThrAlaSerSerThrThrSerTyrValAlaAspGlnThrVal 429
|||||
Db 7041 GAGGACCGCATTGCCCGGTGCGCTCGACAGCCGTTACTGCTGCCCTTTCGCGC----- 6991
QY 430 LysLeuAlaLysThrValLysAspMetSerGlyGluAlaIleSerSerThrGlyAlaSer 449
|||||
Db 6990 ---ATCGTGTCAAGCGCGGAAGAAGCATCGCGGAGCGCTCGATAGCCGACCTGGCA 6934
QY 450 LeuArgSerThrValAsnLeuValArgHisArgSerAlaProGluAlaAspIleGluGlu 469
|||||
Db 6933 CTCGACATGACAATTTCGCGTGTGGAAGAGCGGATCGCGGATGATGATGCGAGAGCC 6874
QY 470 GlyGlyIleSer 473
|||||
Db 6873 TCGTCGCTTTCG 6862

RESULT 8
AE009096/LOCUS 12020 bp DNA linear BCT 20-DEC-2001
DEFINITION Agrobacterium tumefaciens strain C58 circular chromosome, section 122 of 256 of the complete sequence.
ACCESSION AE009096 AE008688
VERSION AE009096.1 GI:17739754
KEYWORDS

SOURCE
ORGANISM
Agrobacterium tumefaciens str. C58 (U. Washington).
Agrobacterium tumefaciens str. C58 (U. Washington)
Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
Rhizobiaceae; Rhizobium.

REFERENCE
AUTHORS
1 (bases 1 to 12020)
Wood,D.W., Setubal,J.C., Kaul,R., Monks,D., Chen,L., Wood,G.E.,
Chen,Y., Woo,L., Kitajima,J.P., Okura,V.K., Almeida Jr.,N.F.,
Zhou,Y., Bovee Sr.,D., Chapman,P., Clendenning,J., Deatherage,G.,
Gillet,W., Grant,C., Guenther,D., Kutayavin,T., Levy,R., Li,M.,
McClelland,E., Palmieri,A., Raymond,C., Rouse,G.,
Saenphimmachak,C., Wu,Z., Gordon,D., Eisen,J.A., Paulsen,I.,
Karp,P., Romero,P., Zhang,S., Yoo,H., Tao,Y., Biddle,P., Jung,M.,
Krespan,W., Perry,M., Gordon-Kamm,B., Liao,L., Kim,S., Hendrick,C.,
Zhao,Z., Dolan,M., Tingey,S.V., Tomb,J., Gordon,M.P., Olson,M.V.
and Nester,E.W.

TITLE
The genome of the natural genetic engineer Agrobacterium

JOURNAL
MEDLINE
PUBMED
11743193
2 (bases 1 to 12020)
Wood,D.W., Setubal,J.C., Kaul,R., Monks,D., Chen,L., Wood,G.E.,
Chen,Y., Woo,L., Kitajima,J.P., Okura,V.K., Almeida Jr.,N.F.,
Zhou,Y., Bovee Sr.,D., Chapman,P., Clendenning,J., Deatherage,G.,
Gillet,W., Grant,C., Guenther,D., Kutayavin,T., Levy,R., Li,M.,
McClelland,E., Palmieri,A., Raymond,C., Rouse,G.,
Saenphimmachak,C., Wu,Z., Gordon,D., Eisen,J.A., Paulsen,I.,
Karp,P., Romero,P., Zhang,S., Yoo,H., Tao,Y., Biddle,P., Jung,M.,
Krespan,W., Perry,M., Gordon-Kamm,B., Liao,L., Kim,S., Hendrick,C.,
Zhao,Z., Dolan,M., Tingey,S.V., Tomb,J., Gordon,M.P., Olson,M.V.
and Nester,E.W.

REFERENCE
AUTHORS
1 (bases 1 to 12020)
Wood,D.W., Setubal,J.C., Kaul,R., Monks,D., Chen,L., Wood,G.E.,
Chen,Y., Woo,L., Kitajima,J.P., Okura,V.K., Almeida Jr.,N.F.,
Zhou,Y., Bovee Sr.,D., Chapman,P., Clendenning,J., Deatherage,G.,
Gillet,W., Grant,C., Guenther,D., Kutayavin,T., Levy,R., Li,M.,
McClelland,E., Palmieri,A., Raymond,C., Rouse,G.,
Saenphimmachak,C., Wu,Z., Gordon,D., Eisen,J.A., Paulsen,I.,
Karp,P., Romero,P., Zhang,S., Yoo,H., Tao,Y., Biddle,P., Jung,M.,
Krespan,W., Perry,M., Gordon-Kamm,B., Liao,L., Kim,S., Hendrick,C.,
Zhao,Z., Dolan,M., Tingey,S.V., Tomb,J., Gordon,M.P., Olson,M.V.
and Nester,E.W.

REFERENCE
AUTHORS
1 (bases 1 to 12020)
Wood,D.W., Setubal,J.C., Kaul,R., Monks,D., Chen,L., Wood,G.E.,
Chen,Y., Woo,L., Kitajima,J.P., Okura,V.K., Almeida Jr.,N.F.,
Zhou,Y., Bovee Sr.,D., Chapman,P., Clendenning,J., Deatherage,G.,
Gillet,W., Grant,C., Guenther,D., Kutayavin,T., Levy,R., Li,M.,
McClelland,E., Palmieri,A., Raymond,C., Rouse,G.,
Saenphimmachak,C., Wu,Z., Gordon,D., Eisen,J.A., Paulsen,I.,
Karp,P., Romero,P., Zhang,S., Yoo,H., Tao,Y., Biddle,P., Jung,M.,
Krespan,W., Perry,M., Gordon-Kamm,B., Liao,L., Kim,S., Hendrick,C.,
Zhao,Z., Dolan,M., Tingey,S.V., Tomb,J., Gordon,M.P., Olson,M.V.
and Nester,E.W.

TITLE
JOURNAL
Karp, P., Romero, P., Zhang, S., Yoo, H., Tao, Y., Biddle, P., Jung, M., Krespan, W., Perry, M., Gordon-Kamm, B., Liao, L., Kim, S., Hendrick, C., Zhao, Z., Dolan, M., Tingey, S. V., Tomb, J., Gordon, M. P., Olson, M. V. and Nester, E. W.
Direct Submission
Submitted (27-SEP-2001) Department of Microbiology, University of Washington, 1959 NE Pacific Ave, Box 357242, Seattle, WA 98195-7242, USA
FEATURES
source
location/Qualifiers
1. 12020
/organism="Agrobacterium tumefaciens str. C58 (U. Washington)"
/strain="C58"
/db_xref="taxon:180835"
68. 538
/gene="Atu1343"
68. 538
/gene="Atu1343"
/note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAL42349.1"
/db_xref="GI:17739755"
/translation="MKQALYIVDVQSPFNPPASIVAEISALAGTMSIATVERHDESVTPFERQLGWKPGRDVSLVPADRIFIKHGYPPEVIDHLISLKLDRVLVCGIQADTCVLAAGFALFDAGLHPTLPWLTVGSSILDRSGELGARLMKHFGAVLTGPALAG"
complement(566. 1066)
/gene="Atu1344"
complement(566. 1066)
/gene="Atu1344"
/note="putative; ORF located using Glimmer"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAL42350.1"
/db_xref="GI:17739756"
/translation="MHVKTLALLALLATVLPARAQTOAIDTLPVSAIFVSSGME DRLPEVKGPDGQLRPPASPTRGYKVIARQGDGTAKIYLIQRIVFTADGPSLENT ELEEFQOMKSYVTIDVRPSSNGASASGPFVTYVLTDPMAKEADSWTILIDELGEMK VEKASN"
complement(1086. 1859)
/gene="Atu1345"
complement(1086. 1859)
/gene="Atu1345"
/note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAL42351.1"
/db_xref="GI:17739757"
/translation="MSHSTYSLSFVASPSEEAQTALKEIKAVYSDTPFEEADVIYALG GDGFMLOILNETMNSGRKYGMNRGSVGFLMNDFRVEGLIQRIAVASGNDFHLMTT TDSGDDEFTALAMNEVSLFRQSHQAALRVEVDGKVRLEELICDGMVATPAGSTAYN FSAHGPIPLPLESPLLALTPYSAFRPRRMGALLPNKYVYDIHVLERDKRPVNAVADHT EYKSVRHVRIAQSODRTAKILSDPDRSMSDRVLAEQFNN"
2164. 3081
/gene="Atu1346"
2164. 3081
/gene="Atu1346"
/note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="phenazine antibiotic biosynthesis related protein"
/protein_id="AAL42352.1"
/db_xref="GI:17739758"
/translation="MALNYDIYDVFTPEAKMAGNPLAVYDADDLSDTMOAIAREMNL SETVFINRSTNPSHAASLRIETPSGELPFAGHPYVGAAYAIAERNRGEADDIDWCY LEEKVGPVRCAYKMRAGAVSFAEFDLPKSSSRVELPLNHAALADALGVSSEHGLFENH

gene
CDS
VPSIWTAGVPFLVPMHNIAAVRDMDFDANLWLRTAPLVEGLLAAYIYCRGVNHA KPHARMFSPEMGIAEDPAGSAAALSGAIIHHFDGLDGHYPLLIHQGVEMERSHIH LRMDIKDNEIARARIGHAVRVASGTFEL"
3303. 5252
/gene="Atu1347"
3303. 5252
/gene="Atu1347"
/note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="tetracycline resistance protein, tetM/teto subfamily"
/protein_id="AAL42353.1"
/db_xref="GI:17739759"
/translation="MRTNLGILAHVDAGKTSLTERLLFDVGVIDKIGSVDTGNTQTD SLELRQGITIRAAVVSFTIGDTVNLIDTPGHPDFIAEVERVLGLDAAVYVSAY EGVQOTRVLVRLALORLAVPELFINKVDYRGVAYRDLADOLRVRPVWSKITG AGSKHVQVAATDTECEPLFTLLCETLAENDELRLDYLLTPDRVDAGRLSLIGQOTA CGVWPAFAGVAMTGAGLPALIAITEMLPARDPDEGISEGKIFKIERGWGKEKISY LHLASGVRLRHILPLPGDPARLTGIQLFKDQRQVANSGLAQIARVTLGAGARVD SVGVDAVADGAFHFAPPTLETRVRSRPSSEHAALNALNQMEODPLGLRNDETNE VFVSLYGEVQKIIQSELSTGFGIEAEFEDSTVICYERLSGGQLOTIFREPNPFA TIGLRVEPRPDGAGNSFALEVEFGOMPASFYRAVEETVEDTLKQGVQWQDCHYAM TAARHTSPSSVIADEFRKLPVWLGTALQAOQFVCEPFDHFHIEAPAAITRLISLA KAGARTKNSVISGVAFTBGTIASAAVQGVQGVPLTSLGLGMETSFSHYQAESPSPRRRSGDPFNECDYLLMRNNAE"
complement(5345. 11812)
/gene="Atu1348"
complement(5345. 11812)
/gene="Atu1348"
/note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAL42354.1"
/db_xref="GI:17739760"
/translation="MAELISLEKFRMHDGSGPLADRKVNGMKPGVCVTRNRNANN KISDSVDETAFOALDLOGAFEKEPRTKTAKPQDEARVKEASROTAPAPARTAP VEQTPRSPNLEANDGSKRSPAMILKSLEGGSIGGALRNATIMSVIMALGGLTAHLL YGNALMSIGSLADLTALPGMAIVGILVPVMLFPAIMMARADLRNARSMEVYA LRLAEPETVASDRIMSVGQAVRREVSAMNDGIERTARATELETLVHSEVNALERSYA DNELRVRSIVQELTAERDAIVNHAERISSIVGAQEOIKEELISVGEELSMRIATTGE AFASMIDTRSAALEKSRASTEAMGSLIAKTENLOALNSSGSTISNEFDMRLNLT STLDERGEVLLERFAIHASTLDSGVESLNSALAEERQONETLSARSELNRNTERGO QVIGSLDTVDLKI STLTEEKGLSFROSLOSTADDAIMDDLRSGLYEERMOATVQY NSAFDEHVAQFASAFDQRAAGSLDKLMESLARINETYAGGSEALDTILTSGLERIGST MTDOSIALATALGTGOEMLENALESRTQAFDAISDAITGOFAEITDNTSHAKIDTVLAE RSNALFGALSASQDRFDEALASRLAITGVSQGTAEHLAAMDERAAAINSVAVER RLTETLETRAAAITGAVSGIEDRISDTLESRTAALHVDYSGAESRIADTLDGRTAALS SAISGVEERADTMSDRTSLDMTFANVEERLSETIDNRTSALTGIVASAEKTIAGAL DSRTATFEGDVVAGAEFRIAETLTDGRTAALNAVVSAGEERITDVLDSRTMA BEKIAELDTRTALGELVASAETRIAGALDSRTDSKTVVSGAEERTDVLDSRTMA LDMFSGVEEKITDILDGRTAALKSAVAGVEDRIAGALDSRTAALSIVSGAEERTIAE ALDSRTIALDMTISGVEERIAEMDARASSLSLAAGVGORLEATAFTLENALSGHE RLETMLGSQAERIAGSLERNSLIEQSVGAANRIENVVEDGSSRAQTVEEGYSRLE NNLQSHEIRTALDORADLAATLSSATTOMGDMSEQAMMIGTTVASSASMLESL ETQODTLQKAIDGSAATLEARLNSAGDIAVKIGEAREIGGATDALSTRITISIGNV TTRLDGTGARIETSLDALQTRVGGDLANVNNSIEDAGRRFADALBDKTAVFARTSDEA AERITGILDEQTRVADTFENRTSRLAETFDAGTARIDERLGTMDRALTIGLENNRT IEKGASDLAVSLRGAVSATQNIQDEARSSALLAKSGSEFAEQVOAQNEAFTKAIEE RSGEIVTRISDAQTRLGQAAAVQOTFSEAGNIYVKAFAEAVRVRSOVGVISFTLTS VESALDARGESIRSALDNRTRELNSMLASRSALSLRIIEKAKPVVEEYATIGREAAE KIVSAAQOSAELLSQNSNGVMVEQAISDYAAAGTDAASKLVAATROSTDMLSEHT AMADAVEQIDKYARSGSDAANKLVAATROSTDMLSEHTNSMAELIVEQSATFNAAVE RAAQFGAADAELNASATRFSESASQAADMVSSSRLLGKIDRLSISGOTIAOVAG IVGRFEEHSKVLQASSELNNAQSSLVGTLFEERODRLSLVGLVKRSEIEITAMRV VGVVENTLNEAEERSQNVAGNLRDNLQASFSDIGRLDETQORARSAQTMRGALISA GODASRSIESTLSDAQKXDELVNRIRLGVESSLSEVDNLGSASEKSNAAMANKET LRQAVEEAVSRFAGATDEITRRSSHDIRRELDATRAELKGAFDLPEAKESAAAMBRA VSEQIKALQDISQLVGRTHQMEVSEPPARAIAATQAAERVRPQPAAPAAVQQR PAQOPAPALRGTLPLENRQYENRQAPAPAPATNPAGRREGGGWISDLRGASQOE

/note="PYRROLINE-5-CARBOXYLATE REDUCTASE (P5CR) (P5C REDUCTASE)"
/codon_start=1
/transl_table=11
/product="AGR_L_1737p"
/protein_id="AAK89441.1"
/db_xref="GI:15159304"
/translation="MSLPQKIGFIGTGATIDAMVRGLLVKPAVPHVWVSORSADISA
RLAADFPQVVISRDNQIMIVDCDTVMLAIRPQIAEEVIRPLRFRDGQKVISVVAATSR
DALLDWIGADVRLAQAIPLPEVARRKGVTAIYPADTDTAIFNVLGNAVECTTKAED
LIAAASALMATYFGIMQTTTEMLAENGLPEDKARAVTPLFAGLSEVALISGKTDFT
EMSREFATKGGINEQVMQDFDRNGGSNALKEALSVMERIAQ"
complement(951..2549)
/gene="AGR_L_1740"
complement(951..2549)
/note="probable MFS transporter PA2701 {imported} -
Pseudomonas aeruginosa (strain PA01)"
/codon_start=1
/transl_table=11
/product="AGR_L_1740p"
/protein_id="AAK89442.1"
/db_xref="GI:15159305"
/translation="MSAAKSSASSAFPLAQPVFAVLMAATVLGNTGSEMRDVASSWLM
TDLSPAPAAVALVOAGTLPIFLAIPAGVLTIDIRKFLIAVOLLASVNTLMVL
AHTGMLSVSALIGLTLFGIGALALMGPTWQAIYPELVKREDIKSAVALNSGLINARS
IGPAVGIGILAAFGAFTYGADVASYEFVIAALMWPRAKNANNALQENFGAFRAGL
RYTRASRPLHVLLRAAIFFAFASAWALLPLVARQLIGDASYGILLGAVGAGAIA
GALVMPKLRAREPDADGILLGAIVTALVMAGLSFAPBKWAIILFLIGAWTTALT
LNGAAQAIILPNWYRGRGLAVYLTVEFGAMTAGSIGWAVGEAAGVPQTLLIGAAGLEI
AGLIMHRLKLPAGDADMPSPSNWMLPAEPPAHDGRPVLLIETVEKDHSAFLHA
LDELQERRRDGAYGWVTEDSADPQKIVEMWVESWAEHLRQHKRVSNAADALQGV
LAYHSGLEKPVVRHFLTINRPGKA"
complement(2546..2986)
/gene="AGR_L_1741"
complement(2546..2986)
/gene="AGR_L_1741"
/note="hypothetical protein"
/codon_start=1
/transl_table=11
/product="AGR_L_1741p"
/protein_id="AAK89443.1"
/db_xref="GI:15159306"
/translation="MTDPSSTRLRLRGLAEIYAAPLRTVALALCSAYIQGPLTKIF
DEPGAIAEMNHFGLOPAAAFVFIIFELTASAMVTVGLRWGALALSGFTLLATFI
ALRFWEMAPGMDRMMAITNAFFEHGLAGAFVIYAAIDLTGTGK"
complement(3196..5073)
/gene="AGR_L_1745"
complement(3196..5073)
/gene="AGR_L_1745"
/note="hypothetical protein PA2699 {imported} -
Pseudomonas aeruginosa (strain PA01)"
/codon_start=1
/transl_table=11
/product="AGR_L_1745p"
/protein_id="AAK89444.1"
/db_xref="GI:15159307"
/translation="MHPDTILHNGRVTTLDRCNP NATAVA IKDGLFEVEVGSDAEVMAL
AGSETKIIDLKGRVLPGLIDNHTVIRGGLNFNMLRWGVRSLADAMYMKROVAI
TPAPQWVRVVGFTERHQVEKERLPTIEINAVADPTPVFLHLHYDRALLNGALRAVG
YTKDTPNP PGGEITRANGNPDTGILLAKPNAGILYSTLAKGKPLIDYQVNSRHFMR
ELNRLGVTGVIDAGGFQNPDDYEVIQKLSDENQMTVRLAYNLFTQKPKQEKEDFLN
WTSVVKYKQGNDFRHNAGAGEMLVESADEFDQRPDMPMEGELEDVVRILAE
RMPWRLHATYDETISRALDYFEKVNRPDIPLEGLNWFDAETISDRSIDRIALGGGI
ATQHRMAYOGEEYFEYRGGVAGATPIARMLEKGVNSAGTDATRVASYNPWYSLFW
MYTGKTVGMOLYPRANCIDRETALRMWTEKVTWFSNEEGKKRIEKGQPADLVVPA
DYFSCAEDEISFLTSELTMVGKIVYGAGDFRVLDESNIIPAMPDMSPVRRFGCAAW
GEPQAGKNSLRRAITSCGCASDCGVGHGHDHAGAWTSKLPITADLKGFEGALGCSCWA
Y"
complement(5178..5531)
/gene="AGR_L_1747"
complement(5178..5531)
/gene="AGR_L_1747"

/note="hypothetical protein"
/codon_start=1
/transl_table=11
/product="AGR_L_1747p"
/protein_id="AAK89445.1"
/db_xref="GI:15159308"
/translation="MPRRRDQYPVRRGFTHPKVSMKYLLSLGAGLIVGIYSLNV
RSPAPVIALVGLGILVGEQIIPARSITDREPAVSWFNQIKPHMGHMPKGAET
TDLAQTQOPSNRKS"
complement(5542..6192)
/gene="AGR_L_1749"
complement(5542..6192)
/gene="AGR_L_1749"
/note="probable hydrolase PA2698 {imported} - Pseudomonas
aeruginosa (strain PA01)"
/codon_start=1
/transl_table=11
/product="AGR_L_1749p"
/protein_id="AAK89446.1"
/db_xref="GI:15159309"
/translation="WIFIDQOPQMAFCVQSIDRQTLKNNVGLAKAKAIFNPTTIT
VETESFGNTFPELLAVFEPENDILERTSMNSWDDQNVRDALAKNAAGRRKIVSGLW
TEVCNTTFALSALHDVADYEIYMWADASGTSLDHKKYAMD RMVQAGV IPTWQOVL
EMQDWMARKETFDAYTSLVKEHSGAYGMGIDYAVTHVHGAEVRKHGRIGPNPAK"
complement(6370..7305)
/gene="AGR_L_1752"
complement(6370..7305)
/gene="AGR_L_1752"
/note="ABC transporter homolog ydfO - Bacillus subtilis"
/codon_start=1
/transl_table=11
/product="AGR_L_1752p"
/protein_id="AAK89447.1"
/db_xref="GI:15159310"
/translation="MMIKDIKGLHVTSMASDARQNNSTFTDTLGLRRYKKTVNFDEP
SVYHLYGDEMGTPGTVMTYFPFDHMMAGRPVGEVGETQFSIPKGAMGEWKDRLIAM
GAAGIQADTVFGANRLRFTGPDGDGLALITANDGRAAWQAEGISSENAIRGFSGARF
NLNDTATEELRLREMGVERADAEGNVTFRVYIPGNGADTIDLATLPNTPEAROGAGSV
HHIAFAVENREKQLEVRKALMDTGQVTPYIDRDYFWAIYFRTPGGILFEIATNEPGF
DRDEDTAHLGEALKLPSRYEPRNQIEANLVSLAA"
7421..8320
/gene="AGR_L_1754"
7421..8320
/gene="AGR_L_1754"
/note="(AF007569) Gstr"
/codon_start=1
/transl_table=11
/product="AGR_L_1754p"
/protein_id="AAK89448.1"
/db_xref="GI:15159311"
/translation="MNDHKALKIFLMAAEKRNFAQVARELSMTAGYTRAIAALECDL
GVQLFVRTTROYSLTDDGAVFAAQIOPAVEALENARDVRNAHKADGRLRISAPTWL
GKTVLPITLSGERELTPKMCFEISLSDGLVNIYVDDYDLAIRISSASDKFTIWRKIR
AVPRILVAAPGSRFTQMKHPSDLTPDDCLAYSGESRRENWILSDGSSYITTAGRAFS
ANNGEVLADMAADGAVAMLPGFHIAEHISGRLVHVLPGWAPPELWLTLYPPYQTL
PPRIATFSKFFEEQLTAHMAVALD"
9016..12045
/gene="AGR_L_1758"
9016..12045
/gene="AGR_L_1758"
/note="1758"
/codon_start=1
/transl_table=11
/product="AGR_L_1758p"
/protein_id="AAK89449.1"
/db_xref="GI:15159312"
/translation="MGSWQOIGVTVIRHSVKDTEMTSILSLSAVQISKISTSTIAGL
TSEDVNALDSTIKIKALSSQIASLSTDNLALFSSEDLRAISVTAVKGLTLTQMAKLSS
AQVSSLSQVTAIYASQIDAMSTDQIKALNSSQVAGLSSAQVATLSDELALFTTDE
IKSISANAIAIGISA AAIAGLSTDNAALTKSI AAMSSTQFNALTSGLATFSADAEVK
AISNKIILAGLDVTKLSTGNIAALSKAQVSAISTQFAAMSTDQIKALTSQVAGLSSA
QVATLSSDELALFSTDEIKAGANAVAGLSAALALTTDNAALFTQIAGLSTQOL
NALTSANLATFSADAEIKAI STKALAGLDVTKLSTGNIAALFKTQAASSLSTQFAAMST

DOI KAL TSEQ VAGLSSA QVATLSSDELALFSTDEISAFSANA VAGISTAA LALTTGN
ATALTKQIAGLSTQNLNLTSGISATFSAD EKAISTKALAGLDVTKLSTGNVALS
KAQVSAISTQFAAMSTDOI KAL TSEQ VAGLSSA QVATLSSDELALFSTDEI KAIGAN
AVAGLSAAALALTTDNASTQI NLTAGMSSTQI NLTANLTAFSAD EKAITTKAL
GGLDVTKLSTGNIAALTKAQVTSLSSTQFALSTDOI KAL TSEQ VSGLSNAQVATLS
DELALFTDEIKSIAANAVAGLSAAALALTTDNASALTFTQIASLSSTQNLNLT SAN
LATFTADEIKAVTTKALAGLDVTKLSTGNVALSKAQVSA LSTQFAAMSTDOI KALT
SDQVAGLSAQVATLSSDELALFSTDEI KAIGATAVSGLSAAALALTTDNASALTFT
QIAGLSSTQNLNLTAFNSANLATFTADEI KAITKALGDLATKLSSTGNIAALSKAQAAAL
STQFALSTDOI KAL TSEQ VAGLSSA QVATLSTDELALFSTDEI KAIGANA IAGLST
AAIGLSSAQAGALS AQQMKVMDAQEAVIKAYKTV"

BASE COUNT 2656 a 3789 c 3495 g 2373 t
ORIGIN

Alignment Scores:

Pred. No.:	15.2	Length:	12313
Score:	149.00	Matches:	127
Percent Similarity:	35.92%	Conservative:	77
Best Local Similarity:	22.36%	Mismatches:	228
Query Match:	6.19%	Indels:	136
DB:	1	Gaps:	25

US-09-825-414-66 (1-487) x AE008285 (1-12313)

QY 7 AlagInGInProProGlyValAlaMetGluSerPheArgThrAlaSerAspAlaSerLeu 26
DB 9367 GCCAGCCAGATCGACGCCATGTCGACCGACAGATCAAGCGCTCAACTCCTTCAGGTC 9426
QY 27 AlaSerSerValArgSerValSerThrThrSerCysArgAspLeuGlnAlaIle--- 45
DB 9427 GCAGGGCTCAGCTCCGCCAGGTCCGCAACGCTGAGCAGCGAAGCTCGCTCTTCACCC 9486
QY 46 ThrAspTyrLeuLysHis-----HisValPheAlaAlaHisArgPheSerValIle 62
DB 9487 ACCGACGAGATCAAGTCGATCAGCGCCCAATGCCATGCCAGTCTTCCGCGCGGCTATC 9546
QY 63 GlySerProAspGluArgAspAlaAlaLeuAlaHisAsnGluGlnIleAspAlaLeuVal 82
DB 9547 GCCGGTCTCAGCACCGACCAATGCAAGCCGACACTGACGAAGTCTCAGATCGCCGCATGTC 9606
QY 83 GluThrArgAlaAsnArgLeuTyrSerGluGluThrProAlaThrIleAlaGluThr 102
DB 9607 TCCACGCAGTTCACACGCCGCTCAGTCGGGCAGCCTTGCACACTTCTCGCGGACGAACTC 9666
QY 103 PheAlaLysAlaGluLysPhe-----AspArgLeuAlaThrThrAla 116
DB 9667 AAGGCGATCAGCAACAGATCCTCGCTGGCTCGATGTCAACCAACTCAGCACCGGCAAC 9726
QY 117 SerSerAlaPheGluAsnThrProPheAlaAlaAlaSerValLeuGlnTyrMetGlnPro 136
DB 9727 ATCGCAGCCCTCAGCAAGGCAAGTTTCGGCGCTGTGACGACACAGTTCGCTGCATG 9786
QY 137 AlaIleAsnLysGlyAspTyrPheAlaThrProLeuLysProLeuThrProLeuIleSer 156
DB 9787 TCGACCGACAG-----ATCAAGGCCCTTCACCTCCGATCAGGTC 9825
QY 157 GlyAlaLeuSerGlyAlaMetAspGlnValGlyThrLysMetMetAspArgAlaArgGly 176
DB 9826 GCCGGTCTCAGCTCCGCG-----CAGGTTGCGAGCCTCAGCAGCGAC----- 9867
QY 177 AspLeuHisTyrLeuSerThrSerProAspLysLeuHisAspAlaMetAlaValSerVal 196
DB 9868 GAACCTGGCTCTTCTTCACCGATGAGATCAAGGCGATCGCGCCCAACGCGGTGCC--- 9924
QY 197 LysArgHisSerProAlaLeuGlyArgGlnValAlaAspMetGlyIleAla----- 213
DB 9925 GGCCTGTCTGCGCAGCCCTGGCGGCTTACCACCGACACAGCTGCAGCACTGACGAAG 9984
QY 214 -----ValGln 215
DB 9985 ACCCAGATCGCGGTCTGTCTTCACACAGCTGAATGCACTGACCTCGCCCAACCTTGCA 10044
QY 216 ThrPheSerAlaLeuAsnValValArgThrValLeuAlaProAlaLeuAlaSerArgPro 235

DB 10045 ACCTTCTCTGCA---GATGAGATCAAGGCCATCAGACCAAGCGCTGGCC----- 10092
QY 236 SerValGlnGlyAlaValAspPheGlyValSerThrAlaGlyGlyLeuValAla----- 253
DB 10093 -----GGTCTCGATGTTACCAAGCTCAGCACCGGCAACATCGCGCGCTGACC 10140
QY 254 -----AsnAlaGlyPheGlyAspArgMetLeuSerValGlnSerArgAspGlnLeu 270
DB 10141 AAAACCCAGGCGGCATCGCTGTCTCAACACAGTTCGACGCCATGTGACCGACAGATC 10200
QY 271 ArgGly-----GlyAlaPheValLeuGlyMetLysAspLysGluProLysAlaAlaLeu 288
DB 10201 AAGGCCCTCACATCCGACGAGGTCCGCCGCTCAGCTCCGCCAGGTTGCGACCTCAGC 10260
QY 289 SerGluGlu-----ThrAspTyrPheAspAlaTyrLysAlaIleLysSer 303
DB 10261 AGCGACGAACTGGCTGTCTTCCACCGATGAGATCAGCGCCTTCAGCGCCAAATGCGGTC 10320
QY 304 AlaSerTyrSerGlyAlaAlaLeuAsnAla----- 313
DB 10321 GCTGGCCTTTCACAGCAGCCCTGGCTGCCCTCACCAACCGGCAATGCACACCGCACTGACG 10380
QY 314 GlyLysArgMetAlaGlyLeu-----ProLeuAspValAlaThrAspGlyLeuLys 330
DB 10381 AAGACACAGATCGCTGGTGTCTTCGACACAGCTCAATGCCCTGACATCAGGCACATC 10440
QY 331 AlaValArgSerLeuValSerAlaThrSerLeuThrLysAsnGlyLeuAlaLeuAlaGly 350
DB 10441 GCCACTTCTCCGCCGATGAAGTCAAGGCCATCAGCACCAAG-----GCACGTGCGCGC 10494
QY 351 GlyTyrAlaGlyValSerLysLeuGlnLysMetAlaThrLysAsnIleThrAspSerAla 370
DB 10495 -----CTCGATGTGACCAAGCTCAGCACCGGCAACGTCGCA--GCCCTC 10536
QY 371 ThrLysAlaAlaValSerGlnLeuSerAsnLeuValGlySerValGlyValPheAlaGly 390
DB 10537 AGCAAGGCACAGTTTCGGCGCTGTGACGACA-----CAGTTCGCTGCC 10581
QY 391 TrpThrThr-----AlaGlyLeuAlaThrAsp 399
DB 10582 ATGTGACCGACAGATCAAGGCCCTGACATCCGACGAGTCCGCGCTCAGCTCCGCG 10641
QY 400 ProAlaValLysLysAlaGluSer-----PheIleGlnAspLysValLysSer 415
DB 10642 CAAGTTGCGAGCTCAGCAGCGACGAACTGGCTCTCTTCTCCACCGATGAGATCAAGCG 10701
QY 416 -----ThrAlaSerSerThrThrSerTyrValAlaAsp 426
DB 10702 ATCGGGCCCAACGCCGTGCCGCGCTGTCTGCCGACGCCCTGGCGGCGCTTACCACCGAC 10761
QY 427 GlnThrValLysLeuAlaLysThr---ValLysAspMetSerGlyGluAlaIleSerSer 445
DB 10762 AATGCATCCGCACTGACGAAGACCCAGATCGCGGCAATGTCTCCACGACGATCAATGCC 10821
QY 446 ThrGlyAlaSerLeuArgSerThrValAsnAsnLeuArgHisArgSerAlaProGlu--- 464
DB 10822 CTG-----ACCTCCGCAAAACCTTGCCACCTTCTCTGACAGAGATC 10863
QY 465 -----AlaAspIleGluGluGly 471
DB 10864 AAGGCAATCACCAACCAAGCTCTCGCGGCTCTGATGTCAACCAAGCTCAGCACCGGCAAC 10923
QY 472 IleSerAlaPheSerArgSerGlu 479
DB 10924 ATCGCGCGCTGACCAAGCGCAG 10947
RESULT 10
AE009328/c 12668 bp DNA linear BCT 20-DEC-2001
LOCUS AE009328
DEFINITION Agrobacterium tumefaciens str. C58 linear chromosome, section 98 of
187 of the complete sequence.
ACCESSION AE009328 AE008689

VERSION AE009328.1 GI:17742413
KEYWORDS
SOURCE Agrobacterium tumefaciens str. C58 (U. Washington).
ORGANISM Agrobacterium tumefaciens str. C58 (U. Washington)
Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
Rhizobiaceae; Rhizobium.
REFERENCE 1 (bases 1 to 12668)
AUTHORS Wood,D.W., Setubal,J.C., Kaul,R., Monks,D., Chen,L., Wood,G.E.,
Chen,Y., Woo,L., Kitajima,J.P., Okura,V.K., Almeida Jr.,N.F.,
Zhou,Y., Boyee Sr.,D., Chapman,P., Clendenning,J., Deatherage,G.,
Gillet,W., Grant,C., Guenther,D., Kutayavin,T., Levy,R., Li,M.,
McClelland,E., Palmieri,A., Raymond,C., Rouse,G.,
Saenphimmachak,C., Wu,Z., Gordon,D., Eisen,J.A., Paulsen,I.,
Karp,P., Romero,P., Zhang,S., Yoo,H., Tao,Y., Biddle,P., Jung,M.,
Krespan,W., Perry,M., Gordon-Kamm,B., Liao,L., Kim,S., Hendrick,C.,
Zhao,Z., Dolan,M., Tingey,S.V., Tomb,J., Gordon,M.P., Olson,M.V.
and Nester,E.W.
TITLE The genome of the natural genetic engineer Agrobacterium
tumefaciens C58
JOURNAL Science 294 (5550), 2317-2323 (2001)
MEDLINE 21608550
PUBMED 11743193
REFERENCE 2 (bases 1 to 12668)
AUTHORS Wood,D.W., Setubal,J.C., Kaul,R., Monks,D., Chen,L., Wood,G.E.,
Chen,Y., Woo,L., Kitajima,J.P., Okura,V.K., Almeida Jr.,N.F.,
Zhou,Y., Boyee Sr.,D., Chapman,P., Clendenning,J., Deatherage,G.,
Gillet,W., Grant,C., Guenther,D., Kutayavin,T., Levy,R., Li,M.,
McClelland,E., Palmieri,A., Raymond,C., Rouse,G.,
Saenphimmachak,C., Wu,Z., Gordon,D., Eisen,J.A., Paulsen,I.,
Karp,P., Romero,P., Zhang,S., Yoo,H., Tao,Y., Biddle,P., Jung,M.,
Krespan,W., Perry,M., Gordon-Kamm,B., Liao,L., Kim,S., Hendrick,C.,
Zhao,Z., Dolan,M., Tingey,S.V., Tomb,J., Gordon,M.P., Olson,M.V.
and Nester,E.W.
TITLE Direct Submission
JOURNAL Submitted (27-SEP-2001) Department of Microbiology, University of
Washington, 1959 NE Pacific Ave, Box 357242, Seattle, WA
98195-7242, USA
FEATURES
source location/Qualifiers
1. 12668
/organism="Agrobacterium tumefaciens str. C58 (U.
Washington)"
/strain="C58"
/db_xref="taxon:180835"
complement(85. 5280)
/gene="ina"
/note="Atu3974"
complement(85. 5280)
/gene="ina"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer"
/transl_table=11
/product="ice nucleation protein homolog"
/protein_id="AAL44776.1"
/db_xref="GI:17742414"
/translation="MTSIIASLVNQIRYLSTEAVAAMTTEDVASLSTQIKALSSDQ
IAALDVEDVKILNSQGLSAISOGAIVNGITLDQNLIDQYAIKSLASASQVSALTTOFH
ALTDDAEALTSQVRLSSAQALASADIAETSTADMAITAKAMPGLGTEVLAGL
TPDQIALNISAISLSTEQIAILSPQAEALTPAQVRLSSLQALAGTDDIATFSA
ADIAITGKAITGLSIEAIELSDRVAALNASAIAGLTTEQIEALSTAOIAALTTAO
IALKTHVAALSTSQVEALSPQVRLSTFASQALVLSADIAETSTADIAAITGTAMP
RLSTEIGTLTLDQIAALTTAOIALNSAQFQALSGOVEALSTHIALSTGVIASL
TADQIEAFSTRQVEALSAAQVRLNSAQIALSPDIAETSTADIAAIAAGKAISGIST
GAIAALTTAOIAALTGAIASLTDQIEALSTQVAALSPQVKSLSSTQIALSVED
LATFSSADLAAITGKAMSGNTDVIALTAALTAQIAAFGASAIAGLTSQVEALSGQVS
LSTTQIKALSTQVAALSGDVAAALSTQFVLSAASIKGLTPEQIAILRPDQFSALT
TAQIAALSSGOITAMGASNTASMAQIAIAISTSSMTGSLSTGOIAALSTEQISRSSR
OMAMLSAQIAALGTEAVAAALSTAOISGLSAGISGLSKQOVATFSTAQVEALSTAOI
LNFSSMIIAELGTEDLAKETTKDIAIAISSSAIAGLSAEIASLTTAOIAALNTQSTNA
LSTAOIAALTTAOVEALSTQVNALSTQIAALSTDIATFSTKIDIAAINSDAIAGLS
AETIASLATGOIALNVRSAISTQVIALTTAQVEALTTQVGAISSTQIALVISTD
DIATFSTRDMAALGSSAIAAGSKDVTVASLTTAQIAGLSMAGISGLSTGOIAALTGDL
NVLTNTQIAALTSQVAAPDVSDITALSTGOIAALSAGAGLTTDQIAALSTDOYGA

gene
CDS
gene
CDS
gene
CDS
MTSGQIALSAKQIALCTDDIATFSTGDIALSSNAVAGLSRDTVASLTTAOIALS
SAGISGLTGOIAGLSTEQVNLVNTNAQISALTSQVAALLEVTDIASIAQIAIGAA
GAGLTTDQIALSTSQVEALSTAOIALNSKQIALSADDLAIFTTAEAMAIGSGAI
SGLPASTIASLTTAQIAALGAAVAGSLTTDQIALGTGOVDALTTAQIAGLSTQVTA
LSVSGISLSAQIAALSTAGVAGLTTDQIALSTSQVEALTSVOIAALSSKQIAALG
ADDDIFTTAEIASIGSSAVAGLSASTIASLTTAOIAALGTAASGLTTDQIYALGTG
QNLGTSQIAGLTSQVAAALSDATFSPDELAIGAGAIKGLSTNFIALSTAEVAA
EALSTQIAALTSDALOGLSTDMLETFSPDELAIGAGAIKGLSTNFIALSTAEVAA
LSTAGVSGLTSEQVDALGKDRIEALSTQITLASSGLAGTLEDMGTSSGELAIS
STAIRGLSNSVAVLSSESIAALTGTQVASLSTGOIAAMDAAQIGALSTSOVSALSAR
QAAALGADDLTTFSAEQITLSSSAIPGLSTPLAGLTTSOAAAFTPGOIAMTSAQV
TALNSSKPANSSVQETIASFLSTPAAATSSAQDSTGNTSGSVGTSTTAQDSSDVASAIL
SYLDY"
complement(5496. 6653)
/gene="g1f"
/note="Atu3975"
complement(5496. 6653)
/gene="g1f"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="UDP-galactopyranose mutase"
/protein_id="AAL44777.1"
/db_xref="GI:17742415"
/translation="MTADERIVIVGAGLSGAVIGRELALAGHLVDIIDARHIANCH
TERDGETGMVHVGPHTFHTDDREVDVNSQTETMPYKRVKTTSGEYSLPVNL
HTINQFEGKNFRPDEARTFIEDKADKSITDPQTFEEQALRFVGRDLYEAFRGYTEKO
WGCSPTLEIPASILKRLPYRFNYDDNFHFKYQGMESGTYDMEIRILDHPTVTKLGT
RFDRAEPATGHVFGSLDGYDFDEFGRLAYRLDFERTYDGYOCCAVMNYGDVS
VPFTRITEHKHFSPEWEDHAGSVCFREFSREGPDIPYPIRLVEDKADYVARAE
REASVTFVGRGLGTRYRLDMDATIREALDARLYIARSRSEGGSPAFHSPV"
complement(6755. 8800)
/gene="Atu3976"
complement(6755. 8800)
/gene="Atu3976"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAL44778.1"
/db_xref="GI:17742416"
/translation="MTANHVSPSNDAAPASLDLARNQRLSLMDLLSVSEALNAAGQO
AAAEFLYKTWAFNDNPTIHLAYFNVAVMLNQGDRAGAVQAFRACLKANPEFAPGH
INLGRALBEDAGLIGHAVQWMSHYAEVTKNVADEFIGHRMHTLQNGRVLLEGAKLIEA
ETALIQAFELRPELPEAGQWASLRORQCKWPVLAPNSISPRKMLDAMSLSLSCYA
DDPMFOQAKAYRLSKTLIGARPDLRSFPRRSPKQSGTGRLRVGLYSSDLRDHAGVF
ALCEVLELHDKDSLVEFYAYCGNVRGTDSTOERIKAAVHCWRDIGHVDDATAQSIIA
DEIDLIDVNGYTKDARAKIFAYRPAPVIVFCGYPGSMGSPFHOYLISDGYMIIPEN
EIVYSEKVLRIACDQPLDRKRPVAARPSRADYGLPEDAFVYASFNGMOKITENCFARW
MTLISETPGSLWLTLTGDDDVNQRDLAEKSGVASERLVFAPKQNPOHARIGLAD
LFLDTEPYGAHSTASDITSGLPVLTMYGKTFAPRCGSIVTAAGVPEMTCSSPEDYV
ARAIGFERDROSLLEVRESIARQRETSVLRDIPALARLLEELFWOMQEGECGETPVP
DLSNLDIYVEAGAELLENIEEEDQSYRERYLKKLRQWHDYAPIPYDRRLMNPNPN"
complement(9292. 12261)
/gene="ina"
/note="Atu3977"
complement(9292. 12261)
/gene="ina"
/note="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer"
/codon_start=1
/transl_table=11
/product="ice nucleation protein homolog"
/protein_id="AAL44779.1"
/db_xref="GI:17742417"
/translation="MTSILSLSAVQISKSTIAGLTSQVNLSDSTKIKALSSSO
IASLSTDNALFSSQEDRAISVTAVKGLTFQMAKLSAQSLSQVTAALYASQID
AMSTDQIKALNSQVAGLSSQAVATLSDELAFETTDEIKSISANAIAIGISAALAGL
STDNAALTKSQIAAMSSSQFNALTSGLSLTASDEVAIKSNKILAGLDVTKLSTGN
AALSKAQSVALSTQFAMSTQDQIKALTSQVAGLSSQAVATLSDELAFSTDEIKA
IGANAVAGLSAALALTTDNAALTKTQIAGLSTQNLALTANLATFSADIEIKAIS

TKLAGLDVTKISTGNIAALTKTQOASLSTQFAMSTDOI KALTSEQVAGLSAQVA
TLSSDELALFSTDEISAFSANAVAGLSTALALTTGNATALTQTQIAGLSSTQNLAL
TSGSIATFSADEIKAITSTKALAGLDVTKLSTGNVNAALSKAOVSALSTQFAMSTDOI
KALTSEQVAGLSAQVATLSSDELALFSTDEIKAIKANAVAGLSAAIALTTDNASA
LTKTOIAGMSTQINALTSTANLATFSADEIKAITTKALGLAGLDVTKLSTGNIAALT
VTSLSSTQFALSTDOI KALTSEQVSGLSAQVATLSSDELALFTTDEIKAVTTKALAGL
GLSAAIALALTNDNASALTKTOIASLSTQNLALSTANLATFTADEIKAVTTKALAGL
DVTKLSTGNVNAALSKAOVSALSTQFAMSTDOI KALTSDQVAGLSTQFAMSTLSSDEL
ALFSTDEIKAIKATVAVSGLSAAIALALTNDNASALTKTOIAGLSSTQNLAMNSANLAT
FTADEVKAITAKALGLDALTSTGNIAALSKAOVAALSTQFALSTDOI KALTSEQ
VAGLSAQVATLSTDELALFSTDEIKAIKANAIAGLSTALIGGLSSAQAGALSQOMK
VMNDAQVEAVIKAVKTV"

BASE COUNT 2342 a 3472 c 4171 g 2683 t
ORIGIN

Alignment Scores:

Pred. No.: 15.8 Length: 12668
Score: 149.00 Matches: 127
Percent Similarity: 35.92% Conservative: 77
Best Local Similarity: 22.36% Mismatches: 228
Query Match: 6.19% Indels: 136
DB: 1 Gaps: 25

US-09-825-414-66 (1-487) x AE009328 (1-12668)

QY 7 AlAGInGInProProGlyValAlaMetGluSerPheArgThrAlaSerAspAlaSerLeu 26
Db 11970 GCCAGCCAGATCGACGCCATGTCAGCCAGACAGATCAAGCGCGCTCAACTCCTCTCAGGTC 11911
QY 27 AlaSerSerSerValArgSerValSerThrThrSerCysArgAspLeuGlnAlaIle--- 45
Db 11910 GCAGGGCTCAGCTCCGCCAGTCGCAACGCTGAGCAGCAGCACTCGCTCTCTCACC 11851
QY 46 ThrAspTyrLeuLysHis-----HisValPheAlaAlaHisArgPheSerValIle 62
Db 11850 ACCGACGAGATCAAGTCGATCAGCGCCCAATGCCATCGCAGGCTTTCCGGCGGCTATC 11791
QY 63 GlySerProAspGluArgAspAlaAlaLeuAlaHisAsnGluGlnIleAspAlaLeuVal 82
Db 11790 GCCGGTCTCAGCACCGACAATGCAGCCGCACTGACGAAGTCTCAGATCGCCGCATGCT 11731
QY 83 GluThrArgAlaAsnArgLeuTyrSerGluGluGluThrProAlaThrIleAlaGluThr 102
Db 11730 TCCACGCAGTTCACGCGCTCAGCTCGGGCAGCCTTGCAACCTTCTCGGGGAGCAAGTC 11671
QY 103 PheAlaLysAlaGluLysPhe-----AspArgLeuAlaThrThrAla 116
Db 11670 AAGCGATCAGCAACAAGATCTGCGCTGCGCTCGATGTCAACCAACTCAGCACCGGCAAC 11611
QY 117 SerSerAlaPheGluAsnThrProPheAlaAlaAlaSerValLeuGlnTyrMetGlnPro 136
Db 11610 ATCGCAGCCCTCAGCAAGCAGCAGGTTTCGGCGCTGTCGACGACACAGATTCCGTCGCATG 11551
QY 137 AlaIleAsnLysGlyAspTyrPleuAlaThrProLeuLysProLeuThrProLeuIleSer 156
Db 11550 TCGACCGACACAG-----ATCAAGGCCCTCACCCTCCGATCAGAGTC 11512
QY 157 GlyAlaLeuSerGlyAlaMetAspGlnValGlyThrLysMetMetAspArgAlaArgGly 176
Db 11511 GCCGGTCTCAGCTCCGCG-----CAGGTTGCGAGCGCTCAGCAGCGAC----- 11470
QY 177 AspLeuHisTyrLeuSerThrSerProAspLysLeuHisAspAlaMetAlaValSerVal 196
Db 11469 GAACCTGGCTCTCTTCTCCACCGATGAGATCAAGCGGATCGGCGCAACGCCGTTGCC--- 11413
QY 197 LysArgHisSerProAlaLeuGlyArgGlnValValAspMetGlyIleAla----- 213
Db 11412 GGCGTGTCTGCGCAGCGCCCTGGCGGCTTACACCGACAACGCTGCACGACTGACGAAG 11353
QY 214 -----ValGln 215
Db 11352 ACCCAGATCGCGGGTCTGTCTTCGACACAGCTGAATGCACCTCCGCCAACCTTGCA 11293

QY 216 ThrPheSerAlaLeuAsnValValArgThrValLeuAlaProAlaLeuAlaSerArgPro 235
Db 11292 ACCTTCTCTGCA--GATGAGATCAAGGCCATCAGCACCAAGCGCTGGCC----- 11245
QY 236 SerValGlnGlyAlaValAspPheGlyValSerThrAlaGlyGlyLeuValAla----- 253
Db 11244 -----GGTCTCGATGTATCAACAGCTCAGCACCAGCAACATCGCCCGCTGACC 11197
QY 254 -----AsnAlaGlyPheGlyAspArgMetLeuSerValGlnSerArgAspGlnLeu 270
Db 11196 AAAACCCAGGCCGATCGCTGTCTCAACACAGTTCGACAGCATGTCCAGCAGCAGATC 11137
QY 271 ArgGly-----GlyAlaPheValLeuGlyMetLysAspLysGluProLysAlaAlaLeu 288
Db 11136 AAGCCCTCACATCCGAGCAGGTGCGGCTCAGCTCCGCGCAGGTTGCGAGCGTTCAGC 11077
QY 289 SerGluGlu-----ThrAspTyrPleuAspAlaTyrLysAlaIleLysSer 303
Db 11076 AGCGACGAACCTGGCTCTCTTCTCCACCGATGAGATCAGCGGCTTCAGCGCAATGCGGTC 11017
QY 304 AlaSerTyrSerGlyAlaAlaLeuAsnAla----- 313
Db 11016 GCTGGCCTTCCACAGCAGCCCTGGCTGCCCTCACCACCGGCAATGCACCGCAGCTGACG 10957
QY 314 GlyLysArgMetAlaGlyLeu-----ProLeuAspValAlaThrAspGlyLeuLys 330
Db 10956 AAGACACAGATCGCTGTGTCTTCGACACACAGCTCAATGCCCTGACATCAGGACGATC 10897
QY 331 AlaValArgSerLeuValSerAlaThrSerLeuThrLysAsnGlyLeuAlaLeuAlaGly 350
Db 10896 GCCACTTCTCCGCCGATGAAGTCAGGCCATCAGCACCACAG-----GCACCTGGCCGGC 10843
QY 351 GlyTyrAlaGlyValSerLysLeuGlnLysMetAlaThrLysAsnIleThrAspSerAla 370
Db 10842 -----CTCGATGTGACCAAGCTCAGCACCAGCGCAAGCTCGCA--GCCCTC 10801
QY 371 ThrLysAlaAlaValSerGlnLeuSerAsnLeuValGlySerValGlyValPheAlaGly 390
Db 10800 AGCAAGCACAAGTTTCGGCGCTGTCGACGACA-----CAGTTCCGCTGCC 10756
QY 391 TyrThrThr-----AlaGlyLeuAlaThrAsp 399
Db 10755 ATGTGACCGACCAAGATCAAGGCCCTGACATCCGAGCAGGTCGCGGCTCTCAGCTCCGCG 10696
QY 400 ProAlaValLysLysAlaGluSer-----PheIleGlnAspLysValLysSer 415
Db 10695 CAAGTTGGCAGCGCTCAGCAGCGCAACTGGCTCTCTTCTCCACCGATGAGATCAAGCGC 10636
QY 416 -----ThrAlaSerSerThrThrSerTyrValAlaAsp 426
Db 10635 ATCGGGCCCAACGCCGTTGCCGCTGTCTGCGCAGCCCTGCGCGGCTTACCACCGAC 10576
QY 427 GlnThrValLysLeuAlaLysThr---ValLysAspMetSerGlyAlaAlaIleSerSer 445
Db 10575 AATGCATCCGCACTGACGAAGACCAAGATCGCGGCGCATGTGCTCCACGACGATCAATGCC 10516
QY 446 ThrGlyAlaSerLeuArgSerThrValAsnAsnLeuArgHisArgSerAlaProGlu--- 464
Db 10515 CTG-----ACCTCCGCAAAACCTTGCCACCTTCTCTGACAGCAGAGATC 10474
QY 465 -----AlaAspIleGluGlyGly 471
Db 10473 AAGGCAATCACCAAGGCTCTCGCGGCTCTCGATGTACCAAGCTCAGCAGCGGCAAC 10414
QY 472 IleSerAlaPheSerArgSerGlu 479
Db 10413 ATCGCGCGCTGACCAAAAGCGCAG 10390

RESULT 11
AE003482/3 301769 bp DNA linear INV 04-OCT-2000
LOCUS Drosophila melanogaster genomic scaffold 14200001386045 section 16
DEFINITION of 17, complete sequence.

ACCESSION AE003482 AE002584
VERSION AE003482.2 GI:10727313
KEYWORDS HTG.
SOURCE Drosophila melanogaster.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 301769)
AUTHORS Adams,M.D., Celniker,S.E., Holt,R.A., Evans,C.A., Gocayne,J.D.,
Amanatides,P.G., Scherer,S.E., Li,P.W., Hoskins,R.A., Galle,R.F.,
George,R.A., Lewis,S.E., Richards,S., Ashburner,M., Henderson,S.N.,
Sutton,G.G., Wortman,J.R., Yandell,M.D., Zhang,Q., Chen,L.X.,
Brandon,R.C., Rogers,Y.H., Blazej,R.G., Champe,M., Pfeiffer,B.D.,
Wan,K.H., Doyle,C., Baxter,E.G., Helt,G., Nelson,C.R., Gabor
Miklos,G.L., Abri,J.F., Agbayani,A., An,H.J.,
Andrews-Pfankoch,C., Baldwin,D., Ballew,R.M., Basu,A.,
Baxendale,J., Bayraktaroglu,L., Beasley,E.M., Beeson,K.Y.,
Benos,P.V., Berman,B.P., Bhandari,D., Bolshakov,S., Borkova,D.,
Botchan,M.R., Bouck,J., Brokstein,P., Brottier,P., Burtis,K.C.,
Busam,D.A., Butler,H., Cadieu,E., Center,A., Chandra,I.,
Cherry,J.M., Cawley,S., Dahlke,C., Davenport,L.B., Davies,P., de
Pablos,B., Delcher,A., Deng,Z., Mays,A.D., Dew,I., Dietz,S.M.,
Dodson,K., Doup,L.E., Downes,M., Dugan-Rocha,S., Dunkov,B.C.,
Dunn,P., Durbin,K.J., Evangelista,C.C., Ferraz,C., Ferriera,S.,
Fleischmann,W., Folsler,C., Gabrielian,A.E., Garg,N.S.,
Gelbart,W.M., Glasser,K., Glodek,A., Gong,F., Gorrell,J.H., Gu,Z.,
Guan,P., Harris,M., Harris,N.L., Harvey,D., Heiman,T.J.,
Hernandez,J.R., Houck,J., Hostin,D., Houston,K.A., Howland,T.J.,
Wei,M.H., Ibegwam,C., Jalali,M., Kalush,F., Karpen,G.H., Ke,Z.,
Kennison,J.A., Ketchum,K.A., Kimmel,B.E., Kodira,C.D., Kraft,C.,
Kravitz,S., Kulp,D., Lai,Z., Lasko,P., Lei,Y., Levitsky,A.A.,
Li,J., Li,Z., Liang,Y., Lin,X., Liu,X., Mattei,B., McIntosh,T.C.,
McLeod,M.P., McPherson,D., Merkulov,G., Milshina,N.V., Mobarry,C.,
Morris,J., Moshrefi,A., Mount,S.M., Moy,M., Murphy,B., Murphy,L.,
Muzny,D.M., Nelson,D.L., Nelson,S.M., Nelson,K.A., Nixon,K.,
Nusskern,D.R., Pacleb,J.M., Palazzolo,M., Pittman,G.S., Pan,S.,
Pollard,J., Puri,V., Reese,M.G., Reinert,K., Remington,K.,
Saunders,R.D., Scheeler,F., Shen,H., Shue,B.C., Siden-Kiamos,I.,
Simpson,M., Skupski,M.P., Smith,T., Spier,E., Spradling,A.C.,
Stapleton,M., Strong,R., Sun,E., Svirskas,R., Tector,C., Turner,R.,
Venter,E., Wang,A.H., Wang,X., Wang,Z.Y., Wassarman,D.A.,
Weinstock,G.M., Weissenbach,J., Williams,S.M., Woodage,T.,
Worley,K.C., Wu,D., Yang,S., Yao,Q.A., Ye,J., Yeh,R.F.,
Zaveri,J.S., Zhan,M., Zhang,G., Zhao,Q., Zheng,L., Zheng,X.H.,
Zhong,F.N., Zhong,W., Zhou,X., Zhu,S., Zhu,X., Smith,H.O.,
Gibbs,R.A., Myers,E.W., Rubin,G.M. and Venter,J.C.
The genome sequence of Drosophila melanogaster
Science 287 (5461), 2185-2195 (2000)
20196006
TITLE JOURNAL
MEDLINE
PUBMED 10731132
REFERENCE 2 (bases 1 to 301769)
AUTHORS Adams,M.D., Celniker,S.E., Gibbs,R.A., Rubin,G.M. and Venter,C.J.
TITLE Direct Submission
JOURNAL Submitted (21-MAR-2000) Celera Genomics, 45 West Gude Drive,
Rockville, MD, USA
COMMENT On Oct 9, 2000 this sequence version replaced gi:7292521.
FEATURES
SOURCE Location/Qualifiers
1. .301769
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/chromosome="3L"
<333. .>3087
/gene="Rpd3"
/note="CG7471"
/map="64B16-64B17"
/db_xref="FLYBASE:FBan0007471"
/db_xref="FLYBASE:FBgn0015805"
join(333. .598,1036. .1148,1227. .2472,2568. .3087)
/gene="Rpd3"
/product="CT1177"
/note="Nucleotide sequence of the Celera sequence differs
from the published sequence for this transcript."

CDS
/db_xref="FLYBASE:FBan0007471"
/db_xref="FLYBASE:FBgn0015805"
join(556. .598,1036. .1148,1227. .2472,2568. .2731)
/gene="Rpd3"
/note="Rpd3 gene product; Nucleotide sequence of the
Celera sequence differs from the published sequence for
this transcript"
/codon_start=1
/protein_id="AAF47924.1"
/db_xref="GI:7292522"
/db_xref="FLYBASE:FBan0007471"
/db_xref="FLYBASE:FBgn0015805"
/translation="MQSHSKRVCYIYDSDIGNYYGQHPMKPHRIRMTNLLNYG
LYRMEIYRPHKATADEMTKFHSDYVRFILRSIRPDNMSYRKQORFNVGEDCPVFD
GLYFCOLSAGGSVAAAVKLKQASEICINWGGGLHAKKSEASGFCYVNDIVLGILE
LKYHQRLYIDIDVHHGDGVEFAFYTTDRVMTVSFHKGXYFPGTGLRDIGAGKGK
YVAVNIPLRDGMDDAYESIIFYPLISKVMTFQPAAVYIQGADSLTGDRLGCFNLTY
KGHGKCYEFVKKYNLPFLWVGSGGYTIRNVSRCWTYETSVALAVEIANELPYNDYFEY
FGPDRKHISPSNMTNQNTSEYLEKIKNLFENLRMLPHAGVQIOAIPEDAINDESD
DEKDKDKDRLQSDKDKRIPENEYSDSEDEGEGRDNRSYKGQRKPRLDKDTNS
NKASSTSSSEIKDEKEKDGADGESTASTNTSNNSNNSNDNAGATANAGSGSGSG
SGAGAKGAKENNI"
complement(<3432. .>3818)
/gene="CG13717"
/evidence=not_experimental
/map="64B16-64B17"
/db_xref="FLYBASE:FBan0013717"
/db_xref="FLYBASE:FBgn0035562"
complement(<3432. .>3818)
/gene="CG13717"
/product="CT33183"
/evidence=not_experimental
/db_xref="FLYBASE:FBan0013717"
/db_xref="FLYBASE:FBgn0035562"
complement(3432. .3818)
/gene="CG13717"
/note="CG13717 gene product"
/codon_start=1
/evidence=not_experimental
/protein_id="AAF47925.1"
/db_xref="GI:7292523"
/db_xref="FLYBASE:FBan0013717"
/db_xref="FLYBASE:FBgn0035562"
/translation="MPNLPQKAQRQPOEHKSRHKATDTEAPKQQQQQQQQQSFYD
ATNANGSSLPASRVKNPEEPGGKSPVPMRQHHSKVTRPPPTLFLENSLQRFQAN
PIYLGGPRLQPRSNRTSIDSLSLD"
complement(<8052. .>8405)
/gene="CG13716"
/evidence=not_experimental
/map="64B17-64B17"
/db_xref="FLYBASE:FBan0013716"
/db_xref="FLYBASE:FBgn0035563"
complement(<8052. .>8405)
/gene="CG13716"
/product="CT33182"
/evidence=not_experimental
/db_xref="FLYBASE:FBan0013716"
/db_xref="FLYBASE:FBgn0035563"
complement(8052. .8405)
/gene="CG13716"
/note="CG13716 gene product"
/codon_start=1
/evidence=not_experimental
/protein_id="AAF47926.1"
/db_xref="GI:7292524"
/db_xref="FLYBASE:FBan0013716"
/db_xref="FLYBASE:FBgn0035563"
/translation="MCLIEQNKVEVKVNSKENPSQNSPPKSPTAAYAMFLHREELRRRK
LHYSRVSVTKIHLTSELIAQTRQNIROCSFEDLEILARETLFKKNLQRLNYRRMHIH
QLMISKKKQQAQDKK"
complement(<11050. .>11361)
/gene="CG13715"
/evidence=not_experimental

/map="64B17-64C1"
 /db_xref="FLYBASE:FBan0013715"
 /db_xref="FLYBASE:FBgn0035564"
 complement(<11050. .>11361)
 /gene="CG13715"
 /product="CT33181"
 /evidence=not_experimental
 /db_xref="FLYBASE:FBan0013715"
 /db_xref="FLYBASE:FBgn0035564"
 complement(11050. .11361)
 /gene="CG13715"
 /note="CG13715 gene product"
 /codon_start=1
 /evidence=not_experimental
 /protein_id="AAF47927.1"
 /db_xref="GI:7292525"
 /db_xref="FLYBASE:FBan0013715"
 /db_xref="FLYBASE:FBgn0035564"
 /translation="MQQLPPQRLSQRCVGIIQQQTSSKRRKPVGSGAGATVVGPA
 TVRRQYSWGPPIVAETDVIDLANSSNGNGNGNGNRKAGKONGYHGVNRRATTTR
 N"
 gene
 complement(<19046. .>49523)
 /gene="axo"
 /note="CG18296"
 /map="64B17-64C1"
 /db_xref="FLYBASE:FBan0018296"
 /db_xref="FLYBASE:FBgn0026729"
 complement(join(19046. .19796,20440. .20558,23933. .24392,
 24553. .24800,29589. .29988,30176. .30390,33096. .33628,
 33693. .34025,37743. .37939,38039. .38561,43730. .44019,
 45440. .45577,46789. .47008,48572. .48845,48998. .49523))
 /gene="axo"
 /product="CT41507"
 /note="Nucleotide sequence of the Celera sequence differs
 from the published sequence for this transcript."
 /db_xref="FLYBASE:FBan0018296"
 /db_xref="FLYBASE:FBgn0026729"
 <21899. .>23751
 /gene="CG7474"
 /map="64B17-64C1"
 /db_xref="FLYBASE:FBan0007474"
 /db_xref="FLYBASE:FBgn0035565"
 join(21899. .22134,22288. .>23751)
 /gene="CG7474"
 /product="CT8014"
 /db_xref="FLYBASE:FBan0007474"
 /db_xref="FLYBASE:FBgn0035565"
 22288. .23751
 /gene="CG7474"
 /note="CG7474 gene product"
 /codon_start=1
 CDS
 mRNA
 gene
 Alignment Scores:
 Pred. No.: 1.56e+03 Length: 301769
 Score: 146.50 Matches: 91
 Percent Similarity: 39.32% Conservative: 60
 Best Local Similarity: 23.70% Mismatches: 156
 Query Match: 6.09% Indels: 77
 DB: 3 Gaps: 13
 US-09-825-414-66 (1-487) x AE003482 (1-301769)
 QY 141 GLYAsPTRpLeuAlaThrProLeuLysProLeuThrProLeuIleSerGIyAlaLeuSer 160
 Db 271261 GCGGCTTAGTGGCGAGGAGCTTCAGACGACGACGCGTCTTGTAGCGGTACGCAATCGCT 271202
 QY 161 GLyAlaMetAspGlnValGIyThrLysMetMetAspArgAlaArgGIyAspLeuHis--- 179
 Db 271201 GGGCAGGACGGTCTGCTGGCGGCACATCAATGATGGCAGAGCGGCTCTCTCCACGGG 271142
 QY 180 -----TyrLeuSerThrSerProAspLysLeu 188
 Db 271141 GGGCAGGTACTGCCGGAAGGACGCTCAGAGACATCAGCGCGGACGAGATCACACGACG 271082

QY 189 HisAspAlaMetAlaValSerVal-----LysArgHisSerProAlaLeuGIy 204
 Db 271081 CACGGTCTGTAGCGGTATCCATCATCAGCCAGGGAGTCTCCACAGCACAGCCTCGGA 271022
 QY 205 ArgGlnValAspMetGIyIleAlaValGln-----ThrPheSer 218
 Db 271021 GCGGACTCCTCAACAGGGGGCAGGTATTTCAGGGGACAGCTCGAGACATCAGCAGCAG 270962
 QY 219 AlaLeuAsnValValArgThrValLeuAlaProAlaLeuAlaSerArgProSerValGln 238
 Db 270961 ACGGATCACACGACGACGGTCTTGTAGCGGTATCCATCATCAGCCAGAGAGTCTCGAC 270902
 QY 239 GLyAlaValAspPheGIyValSerThrAlaGIyGIyLeuValAlaAsnAlaGIyPheGIy 258
 Db 270901 GGC-----AACGGCTTCGGA 270887
 QY 259 AspArgMetLeuSerValGlnSerArgAspGlnLeuArgGIyGIyAlaPheValLeuGIy 278
 Db 270886 GCGCGATTCTC-----AACAGGGGCGAGGTATTCCGGCGACAGCTCGTTTAC--- 270839
 QY 279 MetLysAspLysGIuProLysAlaAlaLeuSerGIuGIuThrAspTrpLeu----- 295
 Db 270838 -----ATCAGCAGCGGCGACGATCACAGCAGGAGTCTGTAGCGGTA 270794
 QY 296 -----AspAlaTyrLysAlaIleLysSerAlaSerTyrSerGIyAlaAlaLeuAsnAla 313
 Db 270793 TCCATCAGCAGCAGCGGGGTCTCAACAGCAACAGCCTCGGAGCAGACTCCTCAACGGG 270734
 QY 314 GLyLysArgMetAlaGIyLeuProLeu--AspValAlaThrAspGIyLeuLysAlaVal 332
 Db 270733 GGGCAGGTACTGCCGGAAGGAAGCTCATGTGACATCAGCAGCAGCAGCGAGTACAGGGCG 270674
 QY 333 ArgSerLeuValSerAlaThrSerLeuThrLysAsnGIyLeu-----AlaLeuAla 349
 Db 270673 GACGGTCTGTACGGTAAACCATCATCATCAGCAAGAGGGGTGTCACAGCAACAGCCTCAGA 270614
 QY 350 GIyGIyTyrAlaGIyValSerLysLeuGlnLysMetAlaThrLysAsnIleThrAspSer 369
 Db 270613 GCGGACTCCTCAACAGGGGGCAGGTATTTCAGGGGACAGCTCGAAMACATCAGCAGCAG 270554
 QY 370 AlaThrLysAlaAlaValSerGlnLeuSerAsnLeuValGIySerValGIy---ValPhe 388
 Db 270553 ACGGATCACACGCGGACAGTCTTGTAGCGGTAGCCATCATCAGCCAGGGAGTCTCGAC 270494
 QY 389 AlaGIyTrpThrThrAlaGIyLeuAlaThrAspProAlaValLysLysAlaGIySerPhe 408
 Db 270493 AGCAACAGCCTCGGAGCGGACTCCTCAACAGGGGCGCAGGTACTCAGGGGAGAGTTTC--- 270437
 QY 409 IleGlnAspLysValLysSerThrAlaSerSerThrThrSerTyrValAlaAspGlnThr 428
 Db 270436 -----GGAGACATCACGACGACGACGATCACACGGCGGAGCTCTT 270395
 QY 429 ValLysLeuAla-LysThrValLysAspMetSer-----GIyGIuAlaIleSe 444
 Db 270394 GTAACGGTAGCCATCATCAGCCAGAGAGTCTCGACGGCAACGGCTTCGGAGCGGATTC 270335
 QY 444 rSerThr-GIyAlaSerLeuArgSerThrValAsnAsnLeuArgHisArgSerAlaProg 464
 Db 270334 CTCACAGGGGCGAGGTATT-----CACCGAGGGGAGCTC 270299
 QY 464 lAlaAspIleGIuGIuGIyLysSerAlaPheSerArgSerGIuThrProPheGlnL 484
 Db 270298 GTTCACATCACGACGCTGGCGCAACTTCAGCGCGCGCACAGGTCTTGTAGCGGTATCCATC 270239
 QY 484 euArgArg 486
 Db 270238 GTCGAGA 270231
 RESULT 12
 SME591788 306250 bp DNA linear BCT 05-JUL-2002
 LOCUS Sinorhizobium meliloti 1021 complete chromosome; segment 7/12.
 DEFINITION

ACCESSION AL591788 AL591688
VERSION AL591788.1 GI:15074579
KEYWORDS
SOURCE
ORGANISM Sinorhizobium meliloti.
Sinorhizobium meliloti
Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
Rhizobiaceae; Sinorhizobium.
REFERENCE 1 (bases 1 to 306250)
AUTHORS Capela,D., Barloy-Hubler,F., Gouzy,J., Bothe,G., Ampe,F., Batut,J.,
Boistard,P., Becker,A., Boutry,M., Cadieu,E., Dreano,S., Gloux,S.,
Godrie,T., Goffeau,A., Kahn,D., Kiss,E., Lelaure,V., Masuy,D.,
Pohl,T., Portetelle,D., Puehler,A., Purnelle,B., Ramsperger,U.,
Renard,C., Thebault,P., Vandenbol,M., Weidner,S. and Galibert,F.
Analysis of the chromosome sequence of the legume symbiont
Sinorhizobium meliloti strain 1021
Proc. Natl. Acad. Sci. U.S.A. 98 (17), 9877-9882 (2001)
MEDLINE 21396507
PUBMED 11481430
REFERENCE 2 (bases 1 to 306250)
AUTHORS Gouzy,J.
TITLE Direct Submission
JOURNAL Submitted (26-JUL-2001) Gouzy J., Submitted on behalf of the MELILO
EU Consortium
COMMENT MELILO EU Consortium:
Laboratoire de Biologie Molculaire des Relations
Plantes-Microorganismes, UMR215-CNRS-INRA, BP27, F-31326 Castanet,
France, Laboratoire de Genetique et Developpement UMR6061-CNRS,
Faculte de Medecine, 2 avenue du Pr. Leon Bernard, F-35043 Rennes,
France, GATC GmbH, Fritz-Arnold-str. 23, D-78467 Konstanz, Germany,
Universitaet Bielefeld, Biologie IV (Genetik) Universitaetstr 25,
D-33615 Bielefeld, Germany, Unite de Biochimie physiologique,
Universite Catholique de Louvain, Place Croix du Sud 2, Bte 20,
B-1348 Louvain-la-Neuve, Belgium, Unite de Microbiologie, Faculte
des Sciences Agronomiques de Gembloux, Avenue Marechal Juin 6,
B-5030 Gembloux, Belgium. E-mail:Jerome.Gouzy@toulouse.inra.fr
http://sequence.toulouse.inra.fr/meliloti.html.
location/Qualifiers
1. .306250
/organism="Sinorhizobium meliloti"
/strain="1021"
/db_xref="taxon:382"
488. .2707
/gene="SMC00954"
488. .2707
/gene="SMC00954"
/function="miscellaneous; hypothetical/global homology"
/note="Product confidence : hypothetical
Gene name confidence : hypothetical
predicted by Codon_usage
predicted by Homology
predicted by Framed"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="CONSERVED HYPOTHETICAL PROTEIN"
/protein_id="CAC46225.1"
/db_xref="GI:15074580"
/db_xref="SPTREMBL:Q92PT4"
/translation="MPAQPHLPANAPNIVIMLDVGPALPDTFGGVIETPTLSRIAQ
DGLSYNRFHNAAMCSPSRALLTGRNHRVGNQJIAELANDWDGYTGRIPTISATYAK
VLGYGYATAAF GKWHNTPANETTAVGPTYDWPAGEGIGDYFYGLAGESSQWEPAY
VENTVRVDPSHGKEGYHTFEDMTNKAVTMKQVHALTPDRPFEMYWAPGAHGPHTIF
KDWADKYKGFDDGWDAMREIRYAROKLWIPONTOLTPRPDTLSGWDGSSAEGQNGTISEL
QRRIMEVPAGYTEHADVOAGRLLEALDELGIRDNLTIFYWGDNGSSAEGQNGTISEL
LAONGIRTEIKDHIRAMNELGMDVLGSPKADNMYHAGNAWAGSTPHRSTKLVAHFG
GTRTPLVSWPGKIKPDQTPRSQFHLNDIVPTIYVDLDIQPKLVDGISQDPLDGS
MTYTFDSPEASGQKEQYFEIMGSKRAIYQDEWIASVFGPRIPWKAGLDPAIFKMSDN
DIWELHDLSDQYSQAKNVADHPEALKRAFGVQAEANKVEPVGGGLWSAVFHPED
APSNPATEFOFTQDVVGVPEFTAPKVGARSNLVTIEALRPDSEGLVALGAFSGGYA
LWYENGKLYEYNLFIEIERPLESSDPLPSGKVNIEVETRKVSSDHAPLDVIRIDG
KEYAKGHVPRSAPLAFETANDAFDVGKDSYSPVALATYFDRKPFPAFNGKIESLKVEYTK"
2887. .4062
/gene="SMC00955"

CDS
2887. .4062
/gene="SMC00955"
/function="miscellaneous; hypothetical/global homology"
/note="Product confidence : hypothetical
Gene name confidence : hypothetical
predicted by Codon_usage
predicted by Framed"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="CONSERVED HYPOTHETICAL PROTEIN"
/protein_id="CAC46226.1"
/db_xref="GI:15074581"
/db_xref="SPTREMBL:Q92PT3"
/translation="MLLYGCAGRPTGVLEPVANPSAAQVEMLVTTTRRAEKGEMF
SGERALSPTEAEMTVSIPPDTVRKAGEVAMPSPKLPNPETDFATLKAVEYDGALEKW
LNTHVRSKPDGSLVVFIFGFNNHFEDAVFRFAOIIDSARSVPVLATMPSRGSLLAY
GYDRESTNYTRNAVERLFOYLARDPEVKEVSIASHSMGNWLAESLRQMAIRNDGLPA
KERNVMLAAPDVVDVFGSOIADMGRKQRPRTLLFVSRDDRALAVSRRWGNVSRLGAI
DPEQSPYKEELIANGITVYDLTKIKAGDNLHHTKFAESPEIVRLIGSRSSGOTLTD
RLGLGDHIVAATAGAAHTVGTAAGLVVAAPAAIVDQNTRRNYVHVEALAEPSGATR"
4548. .4724
/gene="SMC00959"
4548. .4724
/gene="SMC00959"
/function="miscellaneous; unknown"
/note="Product confidence : hypothetical
Gene name confidence : hypothetical
predicted by Codon_usage
predicted by Framed"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="HYPOTHETICAL/UNKNOWN PROTEIN"
/protein_id="CAC46227.1"
/db_xref="GI:15074582"
/db_xref="SPTREMBL:Q92PT2"
/translation="MSKRRLIDTGTDKRYVRDESGKFKHSVDVGRSLSDKKQKAKH
DAKPGQGDKDRKN"
4728. .5522
/gene="xthA1 OR SMC00956"
4728. .5522
/gene="xthA1 OR SMC00956"
/EC_number="3.1.11.2"
/function="macromolecule metabolism; macromolecule
degradation; degradation of dna"
/note="Product confidence : probable
Gene name confidence : probable
predicted by Codon_usage
predicted by Homology
predicted by Framed"
/codon_start=1
/evidence=not_experimental
/transl_table=11
/product="PROBABLE EXODEOXYRIBONUCLEASE III PROTEIN"
/protein_id="CAC46228.1"
/db_xref="GI:15074583"
/db_xref="SPTREMBL:Q92PT1"
/translation="MKIATYNVNGVNGRLGVLLRMLEASPDVVCLOELKADPKFPY
KAIEAGYGAIIWHGQKSWNGVAIILARDREPTLTRKGLPGDDPTHSRYIEAIVEGMVI
GCLYLPNGNPPYGPKPEYKLAWEHRLTAYAAELLEDVPLVILAGDYVMPTLEDVYKP
ERWVNDALFRIEVRDAYHRLLEQGWTDALROLHGERVYTFWDYFRNAPARDAGLRID
HLLSPHVTLRLSAAGYDRHVRGWEHTSDHAPWTIELSDGPAEEDQ"
5519. .5737
/gene="SMC00958"
5519. .5737
/gene="SMC00958"
/function="miscellaneous; unknown"
/note="Product confidence : hypothetical
Gene name confidence : hypothetical
predicted by Codon_usage
predicted by Framed"
/codon_start=1

gene

gene
CDS


```

/evidence=not_experimental
/transl_table=11
/product="HYPOTHETICAL/UNKNOWN PROTEIN"
/protein_id="CAC46229.1"
/db_xref="GI:15074584"
/db_xref="SPTREMBL:Q92PT0"
/translation="MTMNDWRARLRLLEKLVDAVVAGARQRDVYAALIDEIGQLRL
AYERDPPADIDISEQYIEEPPSNWPAAD"
repeat_region
5839. .6005
/note="Sm-4 OR SMC04555
REPEAT SM-4
predicted by Homology"
/evidence=not_experimental
complement(6137. .9652)
/gene="mfd OR SMC04450"
CDS
complement(6137. .9652)

```

```
/note="Product confidence : probable
Gene name confidence : probable
predicted by Codon_usage
predicted by Homology
predicted by Framed"
/codon_start=1
```

```
/evidence=not_experimental
/transl_table=11
/product="PROBABLE TRANSCRIPTION-REPAIR COUPLING FACTOR
(TRCF) PROTEIN"
/protein_id="CAC46230.1"
/db_xref="GI:15074585"
```

/translation-"MTMIPGFDPRKIIATREITIGPVPTGAELVLAELARACAPVA
YILSDGKVADLEQVLGFVAPDIPVTLPLGMDCLPYDRVSPSADTSARRLALASALLA
HRRKHPHAIPLVTINALQVIRISPODYIESLFTARPGAINRMDIDLARBERNGEVRP
TVREMGEEAIVRGILIDVLYVPSGSEPLRIDFEFGDTLPGAINRISFDPASQRITGOVRSIDLNP
PMSEVSLIPETISHFRKQYLSLFGAATRDALQYAVSEGRRYAGMEHMLPLEYDRLET
VFQYLDGERIVTDHLAREAAERSKVLVYDYDARLASASPGKSQYOGTPYKPPVPM
LYTFAGGFEALENDLNAVRLSPFTEHSEGAROVNIEAROGLRWAKAGEADNDGTRT
NVEDQAVKHIAEKRAKAGAKVIVIGCWTESSDLRLQVLAEHGLANIRPYKALSDIGSLK
PGEAASVYLSLEAGFETGDLVIVIGEDTLDRLVRSKRKRKGADFIATFYGLDEGYS
VYHAEHGIGRFVGLRTIEAGAPHPDCELYYADDAKLELPLEVENTIELSRYEGSEGTDAV
YVHAELGVAWQARKAKLKKRLMDAGGLIRIAERHTRHAPVLAQDGYDEFAARFPY
LDLKLGGVAMQARKAKLKKRLMDAGGLIRIAERHTRHAPVLAQDGYDEFAARFPY
DETLEDQNSIDAVRDDLGRGRPMQSLCGDVGFTEVALRAAFIAMDNGOVAVVVP
TTLARQHEFTESDRFGLPIRIRQASBLVGSKDALTKEVAEGKTDIVVGTALLG
SSIKFANQGLLIIDEEOHFQVKKERKELEKTDVHLTISATPIPTLOALITGVREL
SLITTPPYDRAVATFISPFDALVIRETLMREHYRGGSFYCCPRVSDLPEIHDFLKS
DVPELKAAVAGOMPATELEDIMANFAEGRYDYLSTTIVESGLDVPANTLLIVHRAD
MEGLAQYLQGRGVNGSCHKIKEVGFELPYQMLBEAVALKGEETIHGTGMSPOISVGT
HOLDIRAGANLIGEEGSGHIKEVGFELPYQMLBEAVALKGEETIHGTGMSPOISVGT
PVMIPREYVDPDLNRLGLYRRLGELTDLKEIDGFGAELIDRFGLPPEVOHLLKIVV

Alignment Scores:

Pred. No.:	1.71e+03	Length:	306250
Percent:	146.00	Matches:	106
Similarity:	33.93%	Conservative:	84
Best Local Similarity:	18.93%	Mismatches:	208
Query Match:	6.07%	Indels:	162
DB:	1	Gaps:	17

US-09-825-414-66 (1-487) x SME591788 (1-306250)

Qy 3 ILeaSnGLInSerAlaGLInGLInProProGLyValAlaMetCInSerPheArgYThrAlaSer 22
::: ::|||::: ||||| ||| |||
Db 184861 CTGCGCAGCGCGCGCGCTTCGATGCGCCGAGGTTCGATG-----GGCTCGGCCGAG 184911

QY 23 ASPALASerLevalASerSerSerValArgSerValSerThrThrSerCysArgAspLeu 42
::: |||::: ||| ::::|
Db 184912 CCGGAACCAATGCCGCCGACCCGCTCATGTACGGTGGGCCAGGCCGTGCGCCGGAAGTC 184971

QY 43 GlnAlaIeThrAspTyrLeuLysHISHisValPheAlaIaHisArgPheSerValIle 62
|||::: ::::: ::::: ||| ::::

Db 184972 TCGGCATGAACGAGGGCATAGAGCCGACCATCGCCCGGGCAACCGAGCTCGAAGCGCTC 185031

QY	63 glyserproasp	66
----	-----------------	----

Db 185032 GTGCATTTCGGAAGTCAGCGCACTCGAACGCAGCTACAGCGAAACGAACCTGCCGTGTCGCC 185091

QY	67	-----	GLUARGASPLAALALEUALAHISASNGLUGLNILE	78
----	----	-------	-------------------------------------	----

Db 185092 ACGCTGTCAGGAGCTTGGTCTCGAACGCCAAGCGATCATCGGCCATTTCGGATCGGATC 185151

QY 79 AspaLaLeuVal-----Glu 83

Db 185152 CGCAGCGGATCGCTGGCGCGCATACCAAGCTGAAGACGACCTCGAAACGGCAAGCGAG 185211

QY 84 ThrArgAlaAsnArgLeuTyrSerGluGlyGluThrProAlaThrIleAlaGluThrPhe 103

Db 185212 GACATTGGCTGGCGCATCGCCGTCCTCCGGGAGGCCCTTCGCTTCGCTCATCGATACCGG 185271

QY 104 A1A1ySAlAGlulyspheaspargLeuAlaThrThrAlaSerSerAlaPheGluasnThr 123

Db 185272 GCCGCGCA-----CTTACCGACAATCGGACCATGCGCTCGAAAT--- 1853

QY 124 PROPHEAIAAAsERValLeuGlnTYrMetGlnProAlaIleasnLysGlyAspTrp 143

Db 185313 ----- 1853

QY	144	LeuAlaThrProLeuLysProLeuThr---ProLeuIleSerGly-----	157
	: : :	: : :	

Dd 185314 CTGAGCACCATGCTGACCACCCGGACCCGAGCGCTTGTCTTTCGGCCCTGACGACCCGCCGA 185373

QY 158 -----AlaLeuSerCylAlaMetaspGln 165
||||| : : |||

DB 1853/4 G1GGCGCICAGCAACGAATCGACCGCGTCTCGATCGACICAGCGAIAACCTGACCGAG 185433

166 VALGlyThrLysMetMetAspArgALAArgGLYAspLeuHisTyrLeuSerThrSerPro 185

DD 183434 CGTGGCAGCAGTGTCTCAGCCAATTGGAACCCGTGCTTCGACGCTCGACGCCAATACC 1834

```

22 166 asrplysleuhtisaspalametalavalsetvallysarghnlissetptioalaleuglyarg 203
      ::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|

```

UD 183494 GAGAAAGCCTCAACGCCGCCCTGTAACGAACGCCGCCCTCAAGCTCAACGAGACCTGATCCCT 1053

```

27      200 01nvvv1vvvnspmccl0y l1c0n1dvvvd1b1n1m1f1cc0c1m1d1c0u0ns1vva1vvvd1n1g1m1l 220
    ::      |||:::          :::::::::::

```

[illegible]

```

27      220  valDeunnaafIOntadeunnaabeinifiooeetvaldeunnotymdaavaatnspeeneelyval 225
          :::: ::::: ||| ::::||| :::

```

[illegible][illegible][illegible]

100

2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841

[illegible]

C:\Program Files\Easy RM to MP3 Converter\EasyRMtoMP3Converter.exe

DB 105006 ||| ::| ||||| ||| ::::: |||

Our 307 `src[wa]a[atouAcna]ac[wtwAcrcMo+a]ac[wtouBrofouAcnva]a]aThr 336`

[illegible]

On 327 Access[without]Java]Access[without]Auth[without]Access[without] 346

Db 185923 GAC-----AAGATCAATTGCGCTTTCCGAGCAGTCCGCTCACC-----CTG 185964

QY	347	AlaLeuAlaGlyGlyTyrAlaGlyValSerLysLeuGlnLysMetAlaThrLysAsnIle	366
Db	185965	GCAACGGCGCTCGGC-----GCGACGCAGGACTTC	185994
QY	367	ThrAspSerAlaThrLysAlaAlaValSerGlnLeuSerAsnLeuValGlySer-----	384
Db	185995	ATCGAGGAGACGATCGGACGCCGACCTCCGAGCTCAGCAGCCGTGATCGGCACACGCCGAC	186054
QY	385	-----ValGlyValPheAlaGlyTyrThrAla	394
Db	186055	AACCGCATCGAGAGCGTACTTTCGGACAAGACCGGTTGCTGATGGCGCCCTGACCGGAA	186114
QY	395	-----GlyLeuAlaThrAspPro-----AlaVal	402
Db	186115	GCGCAGGAGCGCATCGAAACGGCTTCGGCCAGCGCGCGATGCGCTTGCCAACGCCGCTG	186174
QY	403	LysLysAlaGlySerPheIleGlnAspLysValLysSerThrAlaSerSerThrThrSer	422
Db	186175	ACGACGAGCGAAGCAGCCTGACGAGCAGCAGCTCCCGTACGTCCGCCCTTCATCGAG	186234
QY	423	TyrValAlaAspGlnThrValLysLeuAlaLysThrValLysAspMetSerGlyGluAla	442
Db	186235	GGCCTGCAGTCTGCTCATGCCCCGATCGAGCAGACCCCTTACCGGCTCCACCCGACGAGATC	186294
QY	443	IleSerSerThrGlyAlaSerLeuArgSerThrValAsnAsnLeuArgHisArgSerAla	462
Db	186295	ACGAGCGCGATCGCCCGAGCCAGCATCGCCTCGACAACACGCTTTCGAGAGCGTAGCGCT	186354
RESULT	13		
AE005769/c			
LOCUS	AE005769	13040 bp	DNA linear BCT 12-JUN-2002
DEFINITION	Caulobacter crescentus CB15 section 95 of 359 of the complete genome.		
ACCESSION	AE005769	AE005673	
VERSION	AE005769.1	GI:13422190	
KEYWORDS			
SOURCE	Caulobacter crescentus CB15.		
ORGANISM	Bacteria; Proteobacteria; alpha subdivision; Caulobacter group; Caulobacter.		
REFERENCE	1 (bases 1 to 13040)		
AUTHORS	Niernan,W.C., Feldblyum,T.V., Laub,M.T., Paulsen,I.T., Nelson,K.E., Eisen,J., Heidelberg,J.F., Alley,M.R., Ohta,N., Maddock,J.R., Potocka,I., Nelson,W.C., Newton,A., Stephens,C., Phadke,N.D., Ely,B., Laub,M.T., DeBoy,R.T., Dodson,R.J., Durkin,A.S., Gwinn,M.L., Ely,B., DeBoy,R.T., Dodson,R.J., Durkin,A.S., Gwinn,M.L., Haft,D.H., Kolonay,J.F., Smit,J., Craven,M.B., Khourl,H., Shetty,J., Berry,K., Utterback,T., Tran,K., Wolf,A., Yamathavan,J., Ermolaeva,M., White,O., Salzberg,S.L., Venter,J.C., Shapiro,L. and Fraser,C.M.		
TITLE	Complete genome sequence of Caulobacter crescentus		
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 98 (7), 4136-4141 (2001)		
MEDLINE	21173698		
PUBMED	11259647		
REFERENCE	2 (bases 1 to 13040)		
AUTHORS	Niernan,W.C., Feldblyum,T.V., Paulsen,I.T., Nelson,K.E., Eisen,J., Heidelberg,J.F., Alley,M.R., Ohta,N., Maddock,J.R., Potocka,I., Nelson,W.C., Newton,A., Stephens,C., Phadke,N.D., Ely,B., Laub,M.T., DeBoy,R.T., Dodson,R.J., Durkin,A.S., Gwinn,M.L., Haft,D.H., Kolonay,J.F., Smit,J., Craven,M., Khourl,H., Shetty,J., Berry,K., Utterback,T., Tran,K., Wolf,A., Yamathavan,J., Ermolaeva,M., White,O., Salzberg,S.L., Shapiro,L. and Fraser,C.M.		
TITLE	Direct Submission		
JOURNAL	Submitted (31-JAN-2001) The Institute for Genomic Research, 9712		
FEATURES	Medical Center Dr, Rockville, MD 20850, USA		
source	Location/Qualifiers		
	1..13040		
	/organism="Caulobacter crescentus CB15"		
	/strain="CB15"		
	/db_xref="taxon:190650"		
	complement(260..2020)		
	/gene="CC0921"		
gene	gene		
CDS	CDS		

complement(260..2020)	
/gene="CC0921"	
/note="identified by match to protein family HMM"	
/codon_start=1	
/transl_table=11	
/product="sensor histidine kinase/response regulator"	
/protein_id="AAK22905.1"	
/db_xref="GI:13422191"	
/translation="MGWLAKASDERLDEVAADTYRRLTSRLISALVAVIMFALGP	
AIATGWLISLYIGEGALVTRREKDGRTGTPREAWPAAWASIPINISWTLAILML	
HGEDLMKIAAVAIIWCGVITYQNRHQSALLNLNGLPNASILIEPFEMPDAASAA	
QTARNGILVLVGTLLNVMIQNRAARMDLTRLGREERKALEARAKSTFIATVSH	
ELRTPMNGLLGMAHALQSTIDPVRQEQIALMKSQDSLMQLLNDVLDLSRVETGKVE	
LAPIDMSLDLVGEVVDARDAVFNKLSLSVDYAPDLPPGVHADPLRIQVNTNLS	
NAIKFTVDGVRLEVAARPAANGDQVSITVADTGPVPAAMERVFDSFTQADQTS	
REHGAGIGLSTIARALAROMGDLVLPVSGRACFVFTPAAPACAAPVAAPEEVGANG	
DALGPLQVLMADENAINQLVYRTMLEPLGVALLTVENGQALEAEMERFHCVLMDIN	
MPVMDGITALEAIRDGRGNPAMPVIALTASAMTGDRERFLGLGFDHLGKPKVPVDL	
ITAIKAVNPEPPNAGRRAA"	
2258..5278	
/gene="CC0922"	
2258..5278	
/gene="CC0922"	
/note="identified by match to protein family HMM"	
/codon_start=1	
/transl_table=11	
/product="medium-chain-fatty-acid--CoA ligase"	
/protein_id="AAK22906.1"	
/db_xref="GI:13422192"	
/translation="MQGLMHGALTLDKIIDHAAMHGREVSRVSGPIVRITYGQ	
IRDRAKRVSNALLGMGKSGDRVGTLANWNGRHMFAWYGMIGAVCHTLNPLRPFEG	
IAMIADHAGDRAIFTDLTFLPIIAAILPLRPHVEYVVVPTDRAHMPADFAPAGDAPHF	
KGLICYEELVEQHHADCAWGGTDEGTAAGLCYTSGTTGDPKVMYSHRSNFLHTLTL	
OPVWGLSQKDVIILPVVPMFMANAWGVAFSAPGSGAKMMPGAKMDGASIELDSEQ	
VTFSAVPTVWQMLLOHLBATGATLPTLKKVVIIGGACAPESIIIRAFHDKYDVEVNAW	
GMETSPVGTLSVLDELSKLPYDQOMHWRLKQGRPLGVETICLKDEDEKRLPHDGKA	
EGNLKIRGPITIAEYFRGAGKITLDEGEFDTGDVATIDEGFMQITDRAKDVVKSQG	
EWISTIEIENIANGHPKVALAAVVGQPHPKWDERPVLLVFLKEGETAAPQEFIDYLOG	
KIAKWWMPDDVLIVDDIPLGATGIDKIDKLIRKLRLADGYKLPETPLOSPAPAKPTIA	
AADGSPVKIYAPEPVEVEVAADAVLHTPEAREEETQPAGEPVAASEAISPOTPPEEA	
KLIAAPEVOAKAPEAPAPAPKPVSAADAIIEAATAIAGBPAPAPVTAPEIKAPET	
KASEVKAPPEAKPDAKSAADAPRLRHEPDKAEAPKARPVAEATPFAVPTPGRKKTA	
KASDGKKAKPGWASTYLDFAVYALTPALTAGGALGVKGLIDLPEGYNIMTLDMAP	
RAAQATIGATGLIGLVIALFAGFRMLKAAULLALLITGATPGVMFGAKALFGQAPPIHD	
VATDMKTPLESDAALARGSTALVDDPSLPVGSAPFAGRIAEINAETCPAARPL	
VSSRSPPADAYEAKAAVLASGSLITDDPMDGRLEATASSFWYGLTDDVVVRVRPDA	
GSRLDIRSIGRDAGPDLGRNCDRIGALMTAVKAR"	
complement(5286..6179)	
/gene="CC0923"	
complement(5286..6179)	
/gene="CC0923"	
/note="identified by match to protein family HMM"	
/codon_start=1	
/transl_table=11	
/product="metallo-beta-lactamase family protein"	
/protein_id="AAK22907.1"	
/db_xref="GI:13422193"	
/translation="MIPFVRELEFEYGRCDQVSPILIRRYIARNPGFTFTGTGYIYG	
RGEVAVIDPGLDPEHFEALKAALSGERTVHLVTHNHMDSPLAHPLAQAFGAKYVG	
LPAPSDHGEAAGPSLEEGDDARFRPDVILADGEVLAGPWTLETLATPGHTSNHLCFA	
LREENALFCGDHIMGWSTVITPBDGMDGYFASLAKYARGFETLWPTHGAPVREVA	
PFIDAVVAHRRAREAOILEALGAGPTTIKAMPVSLIYAADVRLHPAAAHSVLAHMIOL	
VREGRVTAGEGALDAEYRLV"	
complement(6513..6920)	
/gene="CC0924"	
complement(6513..6920)	
/gene="CC0924"	
/note="identified by Glimmer2; putative"	
/codon_start=1	
/transl_table=11	
/product="conserved hypothetical protein"	
/protein_id="AAK22908.1"	
/db_xref="GI:13422194"	

gene /translation="MSDFREVTESISVSPQVDPADMARAAAGFVLVINNRPDGEDPG
QPSASVEAARAGMDYLHNVYRGGPTPDQVEIMRAAFEAAGPVLAFCRSGTRSIY
TWSIGQALAGVDRETLVAQGYAAGYDLSGLPA"
7526. .10693
/gene="CC0925"
7526. .10693
/gene="CC0925"

/note="similar to SP:P38370 GB:S64103 PID:402804;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="OmpA-related protein"
/protein_id="AAK22909.1"
/db_xref="GI:13422195"

gene /translation="MNSTKGCYRARLLSTLLAGLASLAAPLITAVATVPTLASAQ
DYSSTGLIGTVDTCNAPVAGAAVQKSLSGFERNVTSQGQFRIALIPSGYSVA
ISKAGFVPASDGNVKGIGASNYGFTIAEGVEVSEIYTGKANPQLDFAGTTGLS
IDVETLSKQVPRIARSVTALTLAPSAVPGDVTGTVNGGSOAAVGAASIGENVEFN
GLNITNFVNGIGATLVPEEFKSVSEIKTGYPAEFGRTGAVINAVTKSGSNDTFAL
HGNEYPNLSRDKSPNTTITQNSLTSRTDRSLIVEAGGPITKDLRFYGLASFQNLETS
TATPGTYYQDRTADPFYAFKLDGYITDKHLELTVEFDSRKRRKRDQTFNPATDVIS
PTRASAESLYOGGENVARYGTFTDMLTVSAAYGRSEVDNAVGNLSGIARVQDTRV
TPYVDRSLQRTAAGTFPSIAERFVRADADLYFKLGSHTIREGYDOENTLTOFTRR
NGGRNITYRRAGAGALGAAVAAQOEYIEGRKFQOTGGFRAGKNKAYIIODAMDNLRL
TSLGVRKDOFOVADPNGOVSSFDNEIGLRGFTYDVFGDRNDKLYGSFGRTYLPVA
SNTAFRAASPATDISEFWLPAGCGTTFGALDAQTGLPTAGLRVLSGSLVACPATVG
SAAPVGAESCVRNNGVSPNPDAVAAKNLSTYQDEYILGEHRENDLMRGVSLIYR
NLGRISEDALLDVGIQNWCAANGYAKTNAARTGCADVWSGEHQYLLINPGSDVKMVL
DPLGETQIRITVSAQQLQFRKAKREYLGLEFSFERTFDGKWGLQGSYTISESKGNF
EGASRSDNGQTDAGITVDFDHLAIFPGTYGLLPNHRGHQLRFGSYAVTDLLGTNL
SVISPRKYGCIGNAPENFDNDGFLBQDYETANASYGAARFCGKVAANRGSVEGEM
VTREPLSARYAVPSRFIPSLPMDKSEGLVLRADVFNVFNLSKIEKDEFEKGTGL
VANPDFQVTAYQPPRYVRFQDWSF"
complement(10690. .11082)
/gene="CC0926"
complement(10690. .11082)
/gene="CC0926"
/note="identified by Glimmer2; putative"

gene /codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAK22910.1"
/db_xref="GI:13422196"
/translation="MGDGARLKRREADAVRPSRLAFGSHLRMGALSLNTPRPHPS
LAGEVGLRSRDEGLKALSGEARETPHPSAFGRHLLPGEEKQIASGISAITVSLIRS
PSLRFLIAKKNGGLVRRARQSPSKRIRP"
complement(11140. .12924)
/gene="CC0927"
complement(11140. .12924)
/gene="CC0927"
/note="identified by match to TIGR protein family HMM
TIGR00705"

CDS /codon_start=1
/transl_table=11
/product="signal peptide peptidase SppA"
/protein_id="AAK22911.1"
/db_xref="GI:13422197"

/translation="MKQFELYAGVFVGLLFFVGVPEFLITAMAGAARPAPVPAHAY
LQDLRNLTDQEPKSPFAEGSSGGSVISIETLRAEKDDVKAILVRLPEGGVAP
AAADELRALFKHREVGKKPIYAHSGLYPSGMVTSTYMLCAATSEFMQPDSSQAV
GISSESMFEKRFEDKYGKAEYQREYKNAVNPYLHSDYTPAHRESTITSMGSVYRT
ALTSAAVDKRDQPTKMLKTELEGPYIAQEAQAKGLIDKYGOVSDIOFALAEKAGAK
LLDFDDYASRSKPTAAVSGPIAVIAGAGALITGTASASPFEGDSNVSDVAQAFR
AATEDKDKAIVERVSSPGSDTASEQILALQAAKKAGRPVVSMTYASGGYMWIS
SQADSIIAQSTLTGSGIGYGGKFAIGEALARFGVDTKDLHVADYSQAFGAGEFTF
EORAKFAGWMDRIYAGFITRVAGEGRLLPPAKVREIAKGRWVTEQAQOLGLVDELGY
YEAVEKAKALAKVKGDIRIKHMOGGASLPFAEFKFLGVSETSAKTMAAAAMVLDGPRS
QALIDEMAKARLRSTPGASVYADMPVR"

BASE COUNT 2203 a 4422 c 4259 g 2156 t
ORIGIN

Alignment Scores: 29.2 Length: 13040
Pred. No.:

Score: 145.00 Matches: 122
Percent Similarity: 36.14% Conservative: 58
Best Local Similarity: 24.50% Mismatches: 178
Query Match: 6.02% Indels: 141
DB: 1 Gaps: 24

US-09-825-414-66 (1-487) x AE005769 (1-13040)

QY 3 ILEASNGINSERIALAGINGLNPROPROGLYVALALAMETGLUSERPHEARGTHRALASER 22
DB 5305 CTGAATACCGCCTCGTGTGAGGACGGCTCAGCGCCCTTGACCGCCCTCATCAGGGCCGCC 5246
QY 23 ASPALASERLEUALASERSESERVALARGSERVALSERTHRTHRSERCYARGSPLEU 42
DB 5245 GATCGCGTCGACGTTGCGGCCAAGATCCGGCCCTGCTCAGCAGCAGATGACCGGATATC 5186
QY 43 GINALAIETHRASPTYRLEULYS-----HISHISVALPHE----- 54
DB 5185 GAGCGCGAACC GGCGCGTCCGGCGCACCCGCCACACACACGTCGTCGTCAGCCGCTA 5126
QY 55 -----ALALAHISARGPHE-SERVALILEGLY-- 63
DB 5125 CCAGAACTGCTGGCGCTGCCCTCCAGCCGACCGCTCCATCGGATCATCGGTGATCAGGCT 5066
QY 64 -SERPROASPGIARGASPALAALALEUALAHISANGLUGLINLEASPALALEUVALGL 83
DB 5065 CATGCCGAGCCGACGAGGCGGCGCTTGCGC-----GCCTCATAGGC 5024
QY 83 UTHRARGALASARGLEUYSERGLUGLYGLUTHRPROALATHRILLEAGLUTHRPH 103
DB 5023 GTCGGCGGGCGAGCGGCTGCTGACACGCGGTCCGCGCGCGGACAGGTCCTCGCGCTTGAT 4964
QY 103 EALA-----LYSALAGLUYSPHEASPARGLEUALATHRTHRAL 116
DB 4963 CTCGGCGATCCGGCGAACC GGCGAAGCGCGCGCGGACCGGAGCGGTAGCGTGCGTCGTC 4904
QY 116 ASERSERALAPHEGLUASNTHRPROPHLEALALA-----ALASE 129
DB 4903 GACCAGCGCC--GCCGTCCCGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4847
QY 129 RVALLEUGLINTYRMETGLNPROALALEASNLYSGLYASPTRPLEUALATHRPROLEULY 149
DB 4846 CGTCTTCACGTGCGTGCGCACGTCGTGATCGCGCGCGCGCTGGCCAAACACAGCGCTTGCC 4787
QY 149 SPRO-----LEUTHRPROLEUILESERGLYALALEUSERGLYALAMETAS 164
DB 4786 CCCGAACATCACGCCAAGGTCGCGCGCGCGCGGTATCAGCAGCCAGAACGCGCGCTTCCA 4727
QY 164 PGLN-----VALGLYTHRYSMETMETASPARGALARGLYASPLE 178
DB 4726 -CAACCGGCCAAGCGCGCGGCAACAGGCGGATGACCAAGCGCGATCAGGCGCGCTCGCATCC 4668
QY 178 WHISTYR--LEUSERTHRSEPROASPLYSLAUHISAPALAMETALA----- 193
DB 4667 CGGCCTGGCG 4608
QY 194 -----VALSERVALYSARGHISERPROALALEUGLYARG----- 205
DB 4607 CGATCAGCGCCACGCTTGACGCCAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4548
QY 206 -----GLNVAL-VALASPMET-----GLYLLEALVALGLNTHR 217
DB 4547 CGAACGAGACCGCAAGTCCAGATAGTGCAGCGCCAGCGCGGCTTGCCCGT 4488
QY 217 HESERALALEUASNVALVALARGTHRVALLEUALPROALALEUALASERARGPROSERV 237
DB 4487 CGCTCGCCTTGCGCGT-----TTGCCCTTGCGGACCGCGGCG 4452
QY 237 AINGINGLYALAVASPHEGLYVALSERTHRVALAGLYGLYLEUVALALASNALAGLYP 257
DB 4451 TGATCGGCACGCGCGAAC--GCCGTGGCCTCGCG-----ACCACGCGCGCGCGCT 4404
QY 257 HEGLYASPARGMETLEUSERVALGLINSERARGASPGINLEUARGGLYALAPHEVAL 277

||||| ||||||| ||| |||||||
Db 4403 TCGGGGCGCTCGGCGCTGTGCGGGTTCGTGACAGA-----CGCAGCGGGCGG----- 4360
Qy 277 euGlyMetLysAsPlySgluProLysAlaAlaLeuSerGluGluThrAspTrpLeuAspa 297
Db 4359 -----TCGGCGCGCG-----GACT 4347
Qy 297 laTyrLysAlaIleLysSerAlaSerTyrSerGlyAlaAlaLeuAsnIaGlyLysArgm 317
Db 4346 TGGCGTCCGGCTTGGCGCTCCGGCGCTTCACCTCGGACGCGCTGGTCTCGGGGGCTTGA 4287
Qy 317 etAlaGlyLeuProLeuAspValaIathraspGlyLeuLysAlaValaArgSerLeuValS 337
Db 4286 TCTCGGGAGCCGTCACCGCGCGCGCGCTTGGGCCCTCGCGCGCGCGATGGCGGCGCTT 4227
Qy 337 eAlaA-----ThrSerLeuThrLysasnGlyLeuAlaLeuAlaGlyLysArgm 355
Db 4226 CGGCGATGATCGCGTCCGGCGCGGAGACCGGCTTGGCCCTCGCGCGCGCGCTCGGGGG 4167
Qy 355 alSerLysLeuGlnLysMetAlaThrLysasnIleThrAspSerAlaThrLysAlaAlav 375
Db 4166 CC-----TTGGCCTGACCTCAGGCTCAGCGCGCA 4137
Qy 375 alSerGlnLeuSerAsnLeuValGlySerVal----- 385
Db 4136 TCAGCTTGGCTTCGGGC-----GGCGCGCTCTGGGGCGAGATGGCCTCCGAGCGCGCA 4083
Qy 386 --GlyValPheAlaGlyTrpThrThrAlaGlyLeuAlaThrAspProAlaValLysLysa 405
Db 4082 CGGGCTCCCGCGCGCGCTGGGTCCTCTTCGCGGGGACGCTCCGGCGTGTGCAAGACAG 4023
Qy 405 laGluSerPheIleGlnAspLysValLysSerThrAlaSerSerThrThr 421
Db 4022 CG-----TCGGCGCGGACCTCTTCGACG 4000
RESULT 14
AF067776 6413 bp DNA linear BCT 07-JAN-1999
LOCUS Abiotrophia defectiva extracellular matrix binding protein (emb)
DEFINITION gene, partial cds.
ACCESSION AF067776 GI:3249002
VERSION AF067776
KEYWORDS Abiotrophia defectiva.
SOURCE Abiotrophia defectiva.
ORGANISM Abiotrophia defectiva.
REFERENCE 1 (bases 1 to 6413)
AUTHORS Manganelli,R. and van de Rijn,I.
TITLE Characterization of emb, a gene encoding the major adhesin of Streptococcus defectivus
JOURNAL Infect. Immun. 67 (1), 50-56 (1999)
MEDLINE 99081722
PUBMED 9864195
REFERENCE 2 (bases 1 to 6413)
AUTHORS Manganelli,R. and van de Rijn,I.
TITLE Direct Submission
JOURNAL Submitted (20-MAY-1998) Microbiology and Immunology, Wake Forest University School of Medicine, Medical Center Boulevard, Winston-Salem, NC 27157, USA
FEATURES
source
1.6413
/organism="Abiotrophia defectiva"
/strain="NVS-47"
/db_xref="taxon:46125"
248..>6413
/gene="emb"
/note="contains 17 tandem repeats of 231bp"
248..>6413
/gene="emb"
/note="adhesin"
/codon_start=1
/transl_table=11
/product="extracellular matrix binding protein"

/protein_id="AAD03320.1"
/db_xref="GI:3249003"
/translation="MLGKNNFHLRLREGEKYTHYAIKLSVGSVAVAGSFLFGSS
NAQAAQVVASQASTSAPEASQGPQEQATEVQASSSQESLESSSEASLATESQA
LSSESSQSSSESSSESSSSSQSSQSETPCKPDETRQTNLKPSSRNDPFRRTPGYAVKP
VEKPATASMRPNNOGQPMPTGTSLRGIGTGKFTDLIEPSGANYAIPVQNLSDNVK
RLUKAVYIRLNPGRGLTEGNIETIDDRGVLITGKRSQDTYKSYLDVYTDLRPATA
EPTVNGTEDKLSSTDATITGTGIPGATIKYTLQDNSVQTRVDRQNWTLINLLKPMR
DSQYITGKGALIEVQIMNGQESAVVRKTYDLGTFLLIENLSSNSRGYKMIAGQK
YIELAVPHDALFMSLEANLTOEGKFGSDNRSQLATRDSITSAMKVGCGTFLRKDT
AKPDLIKKIEVLPGAQFRDYPDTSYHNVDKRIKAGDTVLKKGKDRPLGPELTPV
AALRIVPRYNSGQVVEHRTMSGNATNAPTIAVTTPESTTLNVQDAISDATIKSMRV
NDLEDDARQTNTPYQADIVSYSKNGQVAAIDPTEGEYVTLRARSQDKLSNNTV
QVKVAFTLPREAKNAVNNAKAKNTAIDNNNNULTAEKAEKAEKAEKAKNATLAGID
QAKTTAARNAAQNGCTTDINAVNPVPVAKPANAALAEQAAVKNINEISQRPDLTREEK
QAFMDQVRTARDAMAKVASAANQAVTSARDOGLNAVNNLPTPAKTFEALGHVRQA
ADAKRQAIRDNANLTAEEQADALROVDAAQTPAEBAINONHTNATLAKADSDGVKAIN
DINPQPRSKPAANQALEQVAAAKRQAINNNQULTDEEKAQAIQOVQDALANAKTQVQA
ANDNNGVNOAKTAGTTAINNINPQGTOKAQAIATAIEAAEQAKRLLEQGRDLTTEERN
NALADLTAKAQAKDAVNOARNTGVAGAKDNGVAQIQGINPTAVVKKPARKNAIDQAA
RDKEAEFQANTKLTDEEKAALIKKVQDAPARDAKKAIDRAGSNGDVNNAVQKAAIQA
VDASRSKADVDRAKDOGLQKISDVPAVQPKLNAIAAVDQATDKKAVINNDTLTQE
EKEAAIRKVDDEEAKARQAINDATSNADVAKQOAGTQAINNPQTPAKNAKAAVE
QAADAKQAIENDPNLTROEKDAIAKVDQETNKARQAIIDAATINADVTAKONEGTQA
INAVPQTPAKTDAKNAVTOQAEDKSAIENDPNLTREEKDAKAKVDAEATKAKNAI
DAATSNDEETAKONEGTQAINAVPQTPAKTDAKNAVTOQAADRKDAIENDPNLTREE
KVAAKAKVDAEAKKAKDAIDAATSNADVTAKONEGTQAINDPQTPAKTDAKNAVTO
AADAKDAIEKDPNLTREEKDAKAKVDAEAKKAKDAIDAATSNADVTAKONEGTQAI
NDVPQTPAKTDAKNAVTOQAADAKKDAIEKDPNLTREEKDAKAKVDAEAKKAKDAID
AATSNADVTAKQDAGKNAINAVPQTPAKTDAKNAVTOQAADAKKDAIENDANLTREEK
DAKAKAVDAEATKAKNAIDAATSNADVTAKONEGTKALINDVPQTPAKTDAKNAVDOA
ATTDKSAIENDPALTREEKDAKAKVDAEATKAKNAIDAATSNADVTAKQDAGKNAID
AVPQTPAKTDAKNAVDOAATDKKAIENDPALTREEKDAKAKVDAEAKKAKDAIDA
ATSNADVTAKQDAGKDAINAVPQTPAKTDAKNAVDOAATDKKSAIENDPALTREEK
AVKAKVDAEAKKAKDAIDAATSNADVTAKQTEGTQAINAVPQTPAKTDAKNAVDOAA
TDKKNAIENDPALTREEKDAKAKVDAEAKKAKDAIDAATSNADVTAKQNEGTKAIIND
VPQTPAKTDAKNAVDOAATDKKSAIENDANLTREEKDA"

BASE COUNT 2136 a 1592 c 1598 g 1087 t
ORIGIN
Alignment Scores:
Pred. No.: 14.1 Length: 6413
Score: 143.50 Matches: 123
Percent Similarity: 36.22% Conservative: 78
Best local Similarity: 22.16% Mismatches: 249
Query Match: 5.96% Indels: 105
DB: 1 Gaps: 23
US-09-825-414-66 (1-487) x AF067776 (1-6413)
Qy 3 IleAsnGlnSerAlaGlnGlnPro-----ProGlyValAlaLametGlu 16
Db 2423 ATTAATGAATCAGCCACGCGCCAGATTGACTCGTGAAGAGAACCAAGCCTTCATGGAT 2482
Qy 17 SerPheArgThrAlaSerAspAlaSerLeuAla-----SerSerSerValArgSerVal 34
Db 2483 CAGGTACGGACTGCACGCGATGGCGTATGGCCAAAGTTGCCAGCGCGGCCAACCAATACG 2542
Qy 35 SerThrThrSerCysArgAsp-----LeuGlnAlaIle-----ThrAspTyr 48
Db 2543 GCCGTGACCAAGTCCCGCGACCAAGCGCTGAATGCGGTCAATTAATTACCAACACCAGCA 2602
Qy 49 LeuLysHisHisValPheAlaAlaHisArgPheSerValIleGlySerProAspGluArg 68
Db 2603 GCCAAGTACCCAGAGGCTTTGGGCCAT-----GTGCGCCAGACAGCCGATGCTAAG 2653
Qy 69 AspAlaAlaLeuAlaHisAsn-----GluGlnIleAspAlaLeu---- 81
Db 2654 CGTCAGGCTATTTCGCGACATGCCAACTTAACAGCCGAAGACAAAGCCGATGCCCTGCGT 2713
Qy 82 -----ValGluThrArgAlaAsnArgLeuTyrSerGluGlyGluThrProAla 97

Db 12307 GTGAACGAGGACTTCAGGAG-----AGAACACATCCGAGTGCCAGC 12263
QY 23 AspAlaSerLeuAlaSerSerSerValArgSerValSer-----ThrThrSer 38
Db 12262 TCTGCTACAGACACACATTTACAGATGTGAGCACCAACCCCTCCACCTCTGTGATATG 12203
QY 39 CysArgAspLeuGlnAlaIle-----ThrAspTyrLeuLysHisHisValPhe 54
Db 12202 GTCACCAAGACACACAGCATTCTCTAGCGGACATCCATCTCAAGCCCTCATCAGTTT 12143
QY 55 -----AlaAlaHisArgPheSerValIleGlySerProAspGluArgAspAlaAlaLeu 72
Db 12142 ATAACCTTCTACCAACACATTTACAGATGTGAGCACCAACCCCTCCACCTCTGTGATATG 12083
QY 73 AlaHisAsnGlnGlnIleAspAlaLeuValGluThrArgAlaAsnArgLeuTyrSerGlu 92
Db 12082 ACAGAATCTTCAGGAGTGACCATCACCAACCAAGGTCCTACTGGAGCTGCACACAG 12023
QY 93 Gly-----GluThrProAlaThrIle 99
Db 12022 GGTCCATATCTCTTGGACACATCAACCATGCGCTTACTTGACAGAGACTCCATTAAGCTGTG 11963
QY 100 AlaGluThrPheAlaLysAlaGluLysPheAspArgLeuAla----- 113
Db 11962 ACTCCAGATTTTATGCAATCAGAGAAGACCACTGCATTAAGCAAAAGTCCCAAGATGTG 11903
QY 114 -----ThrThrAlaSerSerAlaPheGluAsn-----ThrProPhe 125
Db 11902 TCCTGGACAAAGCCCTCCCTCTGTGGCAAAACCAAGCTATCCCTCTTCCCTGACACCTTTC 11843
QY 126 AlaAlaAlaSerValLeuGlnTyrMetGlnProAlaIleAsn-----LysGlyAspTyr 143
Db 11842 TTGGTCAACAACCAT-----CCTCCCTGCCACTTCCACGTTACAAAGGCAACAT 11795
QY 144 LeuAlaThrProLeuLysProLeuThrProLeuIleSerGlyAlaLeuSerGlyAlaMet 163
Db 11794 ACATCCTCTCCTGTTCGCGACTTCAGTTCCTTACCCTCGAGCTGTG----- 11747
QY 164 AspGlnValGlyThrLysMetMetAspArgAlaArgGlyAspLeuHisTyrLeuSerThr 183
Db 11746 -----AAGACCACAGATATGTGAACACAAAGCATGGAACCTGTGACCAAT 11702
QY 184 SerProAspLysLeuHisAsp-----AlaMetAlaVal 194
Db 11701 TCACCTCAAAATTTGAACAATCCATCAATGAGATAGTGGCCACTTGGCAGCCACCA 11642
QY 195 SerValLysArgHisSerProAlaLeuGlyArgGlnValValAspMetGlyIleAlaVal 214
Db 11641 GATATGAGACTATTCATCTTCCATAAACAAGCAGTGACCAATATGGG----- 11591
QY 215 GlnThrPheSerAlaLeuAsnValValArgThrValLeuAlaProAlaLeuAlaSerArg 234
Db 11590 ---ACTGCCAGTTTCAGACATGTACTGCATTCACCTC-----CCAGTCAGCTCAGAA 11540
QY 235 ProSerValGlnGlyAlaValAspPheGlyValSerThrAlaGlyGlyLeuValAlaAsn 254
Db 11539 CCATCT-----ACAGCCACATCTCCAATGTTCTGCC 11507
QY 255 AlaGlyPheGlyAspArgMetLeuSerValGlnSerArgAspGlnLeuArgGlyGlyAla 274
Db 11506 TCCAGCATGGGGACGCTCTGCTTATATCA----- 11474
QY 275 PheValLeuGlyMetLysAspLysGluProLysAlaAlaLeuSerGluGluThrAspTyr 294
Db 11473 ---ATACCTGT-----TCTGAGACCACAGAC--- 11450
QY 295 LeuAspAlaTyrLysAlaIleLysSerAlaSerTyrSerGlyAlaAlaLeuAsnAlaGly 314
Db 11450 ----- 11450
QY 315 LysArgMetAlaGlyLeuProLeuAspValAlaThrAspGlyLeuLysAlaValArgSer 334
Db 11449 -----ATTGAGGGAGAGCCACATCCTCCCTGACTGCTGGACGAAAAAGAGACACACC 11396

QY 335 LeuValSerAlaThrSerLeuThrLysAsnGlyLeuAlaLeuAlaGlyTyrAlaGly 354
Db 11395 CTCCAGGAGATGAACTCAACTACAGAGTCAAAACATCATCTCTCCAATGTGTGTGGGG 11336
QY 355 ---ValSerLysLeuGlnLysMet----- 361
Db 11335 GCTATTACTGAAGCCACAATAATGAAAGTCCCTCTTTGTATGCAACATTCATACCACT 11276
QY 362 -----AlaThrLys----- 364
Db 11275 CCTGCTCAGTCAACAAGTTCCAGATATTTCTCAGTAGCCAGAGTAGACTTCAAAC 11216
QY 365 -----AsnIleThrAspSerAlaThrLysAlaAla 374
Db 11215 TCTCCTCCCATGACAATATCTACCCACATGACCACCACAGAGGCTCTTGAGAGCT 11156
QY 375 ValSerGlnLeu-----SerAsnLeuValGlySerValGly----- 386
Db 11155 ACATCAAGATTCCACTTGCCCTTAGACACATCAACCTTGAACCTCAGCAGGACTCCA 11096
QY 387 -----ValPheAlaGlyTyr-----ThrThrAlaGlyLeuAlaThrAsp 399
Db 11095 TCAGTGGTGAAGGAGGTTTGCCCACTCAAAATAAACCACTGCAATGATGATGTC 11036
QY 400 ProAlaValLysLysAlaGluSerPheIleGlnAspLysValLysSerThrAlaSer--- 418
Db 11035 AAGGACGTGTACAGACAAACCCCTCCCTTTCAGATGAAGCCAGCTCTCTCTCAA 10976
QY 419 -----Ser 419
Db 10975 GCACCTGCTGTGTCAACACTTACCTTCTGTGCTTTCACACCCGAATGGCACAGT 10916
QY 420 ThrThrSerTyrValAlaAspGlnThrValLysLeuAlaLysThrValLys----- 436
Db 10915 ACCTCCTCTCCTGTTCTATGTCTCCTCAGTTCCTTACTTCTTCACGTGGTAAAGACCGCAGGC 10856
QY 437 -----AspMetSerGlyGluAlaIleSerSerThrGlyAlaSerLeuArgSerThrVal 454
Db 10855 AAGGTGATACAAGCTTAGAAACAGTGACCAGATTCAACCTCAAGATATGACAAACACTTG 10796
QY 455 AsnAsnLeuArgHisArgSerAlaProGluAlaAspIleGlu 468
Db 10795 GATGACATATCGGTCACTTCAGCAGCCACACAGATATAGAG 10754

Search completed: January 31, 2003, 06:35:09
Job time : 3603.21 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: January 31, 2003, 02:18:04 ; Search time 239.246 Seconds
(without alignments)
4584.089 Million cell updates/sec

Title: US-09-825-414-66

Perfect score: 2407

Sequence: 1 MHINOSAQQPGVAMSFRT.....EEGGISAFSRSETPFQLRRL 487

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 2185239 segs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool/US09825414/runat_28012003_161741_26075/app_query.fasta_1.1294
-DB=N_Geneseq_101002 -QFMT=fastap -SUFFIX=p2n.rng -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=biosum62 -TRANS=human40.cdl
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09825414@cgn_1_1_103_etrunat_28012003_161741_26075 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N_Geneseq_101002.*
1: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	2407	100.0	1464	22	AAD20438	Pseudomonas syring
2	1869.5	77.7	1461	22	AAD20408	P. syringae pv. tom
3	1869.5	77.7	1461	22	AAF55683	Nucleotide sequenc
4	1869.5	77.7	30365	22	AAD20405	P. syringae pv. to
5	148.5	6.2	5397	23	ABL29757	Drosophila melanog
6	147	6.1	4645	24	ABD31881	Lactobacillus rham
7	146.5	6.1	3146	23	ABD23378	Drosophila melanog
8	146.5	6.1	9542	22	AAD04029	Moraxella catartha
9	142	5.9	7029	23	ABD29756	Bacillus subtilis
10	140	5.8	7100	20	AAX24980	Drosophila melanog
11	138	5.7	3690	23	ABD29839	Corynebacterium gl
12	138	5.7	3831	22	AAF67769	Corynebacterium gl
13	138	5.7	6136	22	AAF67767	Drosophila melanog
14	138	5.7	6169	23	ABD29838	C glutamicum codin
15	138	5.7	8979	22	AAH67704	C glutamicum codin
16	138	5.7	349980	22	AAH68532	DNA encoding novel
17	136.5	5.7	4402	23	AAS89717	rsaa gene. Caulob
18	136	5.7	3300	15	AAQ57972	Drosophila melanog
19	135	5.6	2708	23	ABL19629	Shrimp white spot
20	135	5.6	18234	22	AAH62719	Shrimp white spot
21	135	5.6	305107	22	AAH62689	Nucleotide sequenc
22	132.5	5.5	1230025	20	AAX91990	Caulobacter rsaa g
23	132	5.5	3300	17	AAT17717	Caulobacter cresce
24	132	5.5	3300	18	AAV01866	Caulobacter cresce
25	132	5.5	3300	21	AAZ50079	Salmonella secreta
26	131	5.4	1230	18	AAT67029	Staphylococcus aur
27	130.5	5.4	7104	23	AAS51998	Staphylococcus aur
28	130.5	5.4	7107	23	AAS54654	FHA structural gen
29	130	5.4	12036	11	AAQ04668	Staphylococcus aur
30	129.5	5.4	6228	23	AAS52178	Staphylococcus aur
31	129.5	5.4	6561	23	AAS55178	Staphylococcus aur
32	129	5.4	7035	23	AAS54978	Rat glutamate tran
33	129	5.4	7240	22	AAS04273	Drosophila melanog
34	128.5	5.3	3606	23	ABU09625	Staphylococcus pneu
35	128.5	5.3	9345	23	ABU09624	Mycobacterium tube
36	127.5	5.3	7434	23	AAS52179	Fusarium venenatum
37	127.5	5.3	7437	23	AAS55232	B. sphaericus SLP
38	127.5	5.3	4403765	22	AAI99683	Clostridium diffic
39	127	5.3	349980	22	AAH68525	
40	126.5	5.3	2035	23	ABL10547	
41	126.5	5.3	32768	19	AAV52204	
42	126.5	5.3	4411529	22	AAI99682	
43	125.5	5.2	1669	21	AAF08793	
44	125.5	5.2	4197	16	AAQ99430	
45	125.5	5.2	48551	24	AAS20800	

ALIGNMENTS

RESULT 1	
AAD20438	
ID	AAD20438 standard; DNA; 1464 BP.
XX	
AC	AAD20438;
DT	03-JAN-2002 (first entry)
XX	
DE	Pseudomonas syringae pv. tomato strain DC3000 HopPtoA2 DNA.
XX	
KW	Conserved Effector Loc1; CEL; cytosstatic; antibacterial; gene therapy;
KW	Exchangeable Effector Loc1; EEL; disease resistance; transgenic plant;
KW	eukaryotic cell death; cancer; ds.
XX	
OS	Pseudomonas syringae.
XX	
FH	key
FT	CDS Location/Qualifiers
	1..1464

|||||
Db 1201 GCGTTAAGAAAGCCGAGTCGTTTATACAGATGAAGTGAATCGACCGCATCTAGTACC 1260
QY 421 ThrSerTyrValAlaAspGlnThrValLysLeuAlaLysThrValLysAspMetSergly 440
Db 1261 ACAAGCTATGTTGCGACCGACCGCTCAAACTGGCGAAACAGTCAAGGACATGAGCGGG 1320
QY 441 GluAlaIleSerSerThrGlyAlaSerLeuArgSerThrValAsnAsnLeuArgHisArg 460
Db 1321 GAGCGATCTCCAGCACCGGTGCCAGCTTACGCACTACTGTCAATAACCGTCGTCATCCG 1380
QY 461 SerAlaProGluAlaAspIleGluGluGlyIleSerAlaPheSerArgSergluThr 480
Db 1381 TCCGCTCCGGAAGCTGATATCGAAGAAGGTGGGATTTCCGCCGTTTCTCGAAGTGAACA 1440
QY 481 ProPheGlnLeuArgArgLeu 487
Db 1441 CCGTTTCAGCTCAGCGCTTTG 1461
RESULT 2
AAD20408
ID AAD20408 standard; DNA; 1461 BP.
XX
AC AAD20408;
XX
DT 03-JAN-2002 (first entry)
XX
DE P. syringae pv. tomato (Pto) DC3000 CEL open reading frame 5 (ORF5) DNA.
XX
KW Conserved Effector Loci; CEL; cytostatic; antibacterial; gene therapy;
KW Exchangeable Effector Loci; EEL; disease resistance; transgenic plant;
KW eukaryotic cell death; cancer; ds.
XX
OS Pseudomonas syringae.
XX
FH Key Location/Qualifiers
FT CDS 1..1461
FT /*tag= a
FT /product= "Pto DC3000 CEL ORF5 protein"
XX
PN WO200175066-A2.
XX
PD 11-OCT-2001.
XX
PF 03-APR-2001; 2001WO-US10698.
XX
PR 03-APR-2000; 2000US-194160P.
PR 11-AUG-2000; 2000US-224604P.
PR 17-NOV-2000; 2000US-249548P.
XX
PA (CORR) CORNELL RES FOUND INC.
PA (UYNE-) UNIV NEVADA & COMMUNITY COLLEGE SYSTEM.
PA (UYNE-) UNIV NEBRASKA.
XX
PI Collmer A, Alfano JR, Charkowski AO;
XX
DR WPI; 2001-639361/73.
DR P-PSDB; AAE12573.
XX
PT New nucleic acid molecules encoding proteins or polypeptides of
PT Pseudomonas Conserved Effector Loci and Exchangeable Effector Loci
PT genomic sequences, for imparting disease resistance to plants -
XX
PS Claim 1; Page 21; 217pp; English.
XX
CC The invention relates to an isolated nucleic acid molecule comprising a
CC nucleotide sequence encoding proteins or polypeptides of Pseudomonas
CC Conserved Effector Loci (CEL) and Exchangeable Effector Loci (EEL)
CC genomic sequences. CEL and EEL DNA are useful for imparting disease
CC resistance to a plant, by transforming a plant cell with the nucleic acid
CC and regenerating a transgenic plant from the transformed plant cell,
CC where the transgenic plant expresses a heterologous DNA molecule under
CC conditions effective to impart disease resistance, or by treating a plant

CC with an isolated protein or polypeptide, by applying the protein or
CC polypeptide in an isolated form or by applying a non-pathogenic bacteria
CC which secretes the protein or polypeptide, under conditions effective to
CC impart disease resistance to the treated plant. CEL and EEL proteins
CC are useful for causing eukaryotic cell death, by introducing a cytotoxic
CC pseudomonas protein into a eukaryotic cell under conditions effective to
CC cause cell death. CEL and EEL proteins are also useful for treating a
CC cancerous condition, by introducing a cytotoxic Pseudomonas protein into
CC cancer cells of a patient under conditions effective to cause death of
CC cancer cells, and thus treating the cancerous condition. The method
CC further involves administering a targeted DNA delivery system
CC comprising a DNA molecule encoding the cytotoxic Pseudomonas protein,
CC to the patient, where the targeted DNA delivery system delivers the
CC DNA molecule into cancer cells and the cytotoxic Pseudomonas protein
CC is expressed in the cancer cells. The present sequence is
CC Pseudomonas syringae pv. tomato (Pto) DC3000 CEL ORF5 DNA.
XX
SQ Sequence 1461 BP; 311 A; 440 C; 451 G; 259 T; 0 other;
XX
Alignment Scores:
Pred. NO.: 9.59e-148 Length: 1461
Score: 1869.50 Matches: 381
Percent Similarity: 85.19% Conservative: 33
Best Local Similarity: 78.40% Mismatches: 71
Query Match: 77.67% Indels: 1
DB: 22 Gaps: 1
US-09-825-414-66 (1-487) x AAD20408 (1-1461)
QY 1 MethisIleAsnGlnSerAlaGlnGlnProProGlyValAlaMetGluSerPheArgThr 20
Db 1 ATGCACATCAACCGCAGCGCTCCACAAACCCGCTGTGACTGCGACGGATAGCTTTCCGACA 60
QY 21 AlaSerAspAlaSerLeuAlaSerSerSerValArgSerValSerThrThrSercysArg 40
Db 61 GCGTCCGACGCGCTCTTCTTCCCTCCAGCTGTGCGATCTGTCAAGCTCCGATCAGCAACGC 120
QY 41 AspleuGlnAlaIleThrAspTyrLeuLysHisValPheAlaAlaHisArgPheSer 60
Db 121 GAGATAAATGCGATTGCCGATTACCTGACACAGATCATGTGTTGCGTGCATAAACTGCCG 180
QY 61 ValIleGlySerProAspGluArgAspAlaAlaLeuAlaHisAsnGluGlnIleAspAla 80
Db 181 CCGGCCGATTCCGGCTGATGGCCAAAGCTGCAGTTGACGTACACAATGGCGAGATCACTGCG 240
QY 81 LeuValGluThrArgAlaAsnArgLeuTyrSergluGlyGluThrProAlaThrIleAla 100
Db 241 CTGATCGAGACGGCGCCAGCCGCTGCACCTTCGAAGGGGAACCCCGGCAACCATCGCC 300
QY 101 GluThrPheAlaLysAlaGluLysPheAspArgLeuAlaThrThrAlaSerSeraIaPhe 120
Db 301 GACACCTTCGCCAAGGCGGAAAGCTCGACCGATTTGGCGAGCACTACATCAAGCGCGTTG 360
QY 121 GluAsnThrProPheAlaAlaAlaSerValLeuGlnTyrMetGlnProAlaIleAsnLys 140
Db 361 CGGGCGACGCCCTTTCCTCATGGCCTCGTTGCTTCAGTACATGACGCTCGCATCAACAAG 420
QY 141 GlyAspTrpLeuAlaThrProLeuLysProLeuThrProLeuIleSerglyAlaLeuSer 160
Db 421 GCGGATTGGTGGCTCGGCTCCGCTCAAAACCGTGAACCCGCTCATTTCCGGAGCGCTGTCG 480
QY 161 GlyAlaMetAspGlnValGlyThrLysMetMetAspArgAlaArgGlyAspLeuHisTyr 180
Db 481 GCGCGCATGGACAGGTGGGCAACAAGATGATGACCGCGGAGCGGTGATCTGCATTAC 540
QY 181 LeuSerThrSerProAspLysLeuHisAspAlaMetAlaValSerValLysArgHisSer 200
Db 541 CTGAGCGCCTCGCCGACAGGCTCCACGATGCGATGGCCGCTTCGGTGAAGCCCACTCG 600
QY 201 ProAlaLeuGlyArgGlnValValAspMetGlyIleAlaValGlnThrPheSerAlaLeu 220
Db 601 CCAAGCCTTGCTCGACAGGTTCTGTGACACACGGGGGTTGCGGTTCAAGACGTACTCGGCGCG 660

OY	221	AsnValValAlrgrThrValLeuAlaProAlaLeuAlaSerArgrProserValGlnGlyAla	240
Db	661	AACGCCGTACGTACCCTATTTGGCTCCGGCACTGGCGTCCAGACCCGCCGTGCAGGGTGCT	720
OY	241	ValAspPheGlyValSerThrAlaGlyGlyLeuValAlaAsnaIaglyPheGlyAsparg	260
Db	721	GTEGACCTTGCTGTATCATGTGGCGGGTGCTTGCCAAAGCAGGCTTTGGCAACCGC	780
OY	261	MetLeuSerValGlnSerArgaspGlnLeuArgGlyGlyAlaPheValLeuGlyMetLys	280
Db	781	CTGCTCAGTGTGCAGTCGCGGTGATCACCAAGCGTGGCGGTGCATTAGTGCCTCGTTGAAG	840
OY	281	AspLysGluProLysAlaAlaLeuSerGluGluThrAspTrpLeuAspaLatyrLysAla	300
Db	841	GATTAAGAAGCCCCAAGGCTCAACTGAGCGAAGAAAACGACTGGCTCGAGGCTTATAAGCA	900
OY	301	IleLysSerAlaSerTyrSerglyAlaAlaLeuAsnaIaglyLysArgMetaIaglyLeu	320
Db	901	ATCAATCGGCCAGCTACTCGGGTGGCGGCTCAACGCTGGCAAGCGGATGGCCGGTCTG	960
OY	321	ProLeuAspValAlaThrAspGlyLeuLysAlaValArgSerLeuValSerAlaThrSer	340
Db	961	CCACTGATATATGGCGACCGCAATGGGTGGGTAAGAAGCCTGTGTCTCAGCGTCCAGC	1020
OY	341	LeuThrLysAsnglyLeuAlaLeuAlaGlyGlyTyraIaglyValSerLysLeuGlnLys	360
Db	1021	CTGACCCCAAACGGTCTGGCCCTGGCGGTGCTTTGCAGGGGTAGGCAAGTTGCAGGAG	1080
OY	361	MetAlaThrLysAsnIleThrAspSerAlaThrLysAlaAlaValSerGlnLeuSerAsn	380
Db	1081	ATGGCGACGAAAAATATCACCGACCCGGCGAACCAAGCCGGTCACTGATGACCAAC	1140
OY	381	LeuValGlySerValGlyValPheAlaGlyTrpThrThrAlaGlyLeuAlaThrAspro	400
Db	1141	CTGGCAGGTTCCGGCAGCCGTTTTCCGAGGCTGGACCAACGCCGCCGCTGACACAACGATCCC	1200
OY	401	AlaValLysLysAlaGluSerPheIleGlnAspLysValLysSerThrAlaSerSerThr	420
Db	1201	GCGGTGA AAAAGCCGAGTCTTCATACAGGACGCGTGA AATCGACTGCATCCAGTACC	1260
OY	421	ThrSerTyrValAlaAspGlnThrValLysLeuAlaLysThrValLysAspmetSergly	440
Db	1261	ACAGGCTACGTAGCCGACCGACGACCGTCAAACTGGCGGAAGACCGTCA AAGACATGGCGGG	1320
OY	441	GluAlaIleSerSerThrGlyAlaSerLeuArgSerThrValAsnAsnLeuArgHisArg	460
Db	1321	GAGGCGATCACCCATATCCGGCGCAGCTTGGCAATACGGTCAATTAACCTGCGTCAACGC	1380
OY	461	SerAlaProGluAlaAspIleGluGluGlyIleSerAlaPheSerArgSergluThr	480
Db	1381	CCGGCTCGTGAAGCTGATATAGAAGAGGGGGGACGGCGGCT--TCTCAAGTGAATA	1437
OY	481	ProPheGlnLeuArgarg 486	
Db	1438	CCGTTTCGGCCTATGCGG 1455	
RESULT 3			
ID	AAF55683	standard; DNA; 1461 BP.	
XX	AAF55683;		
DT	11-JUN-2001	(first entry)	
DE	Nucleotide sequence of a Hoppta protein of Pseudomonas syringae.		
KW	HIV; tat protein; effector protein; transduction domain; Hoppta protein;;		
OS	Pseudomonas syringae.		
FH	Key	location/Qualifiers	
FT	CDS	1..1461	

```

FT      /*tag= a
FT      /product= "HopTo"
XX      WO200119393-A1.
XX      PD      22-MAR-2001.
XX      PF      13-SEP-2000; 2000WO-US24977.
XX      PR      13-SEP-1999; 99US-0153507.
XX      PA      (CORR ) CORNELL RES FOUND INC.
PI      COLLMER A, Beer SV;
XX      WPI: 2001-257850/26.
DR      P-PSDB; AAB67677.
XX      PT      Delivering effector proteins into target cell for use in protein
PT      therapy, involves introducing effector protein fused to protein
PT      transduction domain of human immunodeficiency virus TAT protein, into
PT      target cells -
XX      PS      Example 2; Page 22-23; 43pp; English.
XX      CC      The present sequence encodes a HopTOA protein of Pseudomonas syringae
CC      pv. tomato DC300 CEL.. This is an effector protein, which is used in
CC      the method of the invention. The specification describes a method for
CC      delivering effector proteins into a target cell. The method comprises
CC      introducing an effector protein fused to a protein transduction domain
CC      of a human immunodeficiency virus (HIV) tat protein into the target
CC      cell. The method is used for delivering a heterologous effector protein
CC      such as an effector protein produced by bacterial plant pathogen, animal
CC      pathogen or a rhizosphere bacteria, or a protein secreted and/or
CC      delivered into eukaryotic cells by a Type III secretion system or a
CC      hypersensitive response elicitor, an avirulence protein, a
CC      hypersensitive response and pathogenicity-dependent outer protein, a
CC      virulence protein or a pathogenicity protein, into an eukaryotic cell.
CC      The method is useful for delivering effector proteins for use in
CC      pharmaceutical, insecticidal, fungicide, herbicide and other
CC      applications.
XX      SQ      Sequence 1461 BP; 311 A; 440 C; 451 G; 259 T; 0 other;

Alignment Scores:
Pred. No.:          9.59e-148          Length:          1461
Score:              1869.50             Matches:           381
Percent Similarity: 85.19%               Conservative:       33
Best Local Similarity: 78.40%            Mismatches:         71
Query Match:        77.67%                Indels:             1
DB:                  22                    Gaps:               1

US-09-825-414-66 (1-487) x AAF55683 (1-1461)
QY      1 MethIstleasnnglnSerAlaglnglnProPoglyValAlametglUserPheargthr 20
QY      |||||||::: ||||| ||| :::|||||
Db      1 ATGCACATCAACCGACGCGCTCCAACAACCGCCTGTGACTGCCAGCATAGCTTTGGACA 60
QY      21 AlaserAspAlaserLeualaSerSerValargSerValSerThrThrSercysarg 40
QY      ||||||| ||||||| ||||||| ||||||| ||||||| |||
Db      61 GCGTCGACGCGTCTTGCCCTCCAGCTCTGTGGCATGTGTCAAGCTCCGATCAGCAACGC 120
QY      41 AspleuglnalaleThrAsptyrLeuLysHisValpheAlalaHisargpheser 60
QY      ::::: ||||| ||||| ||||| ||||| ||||| |||
Db      121 GAGATAAATGGCATTCGCCGATTACCTGCACAGATCATGTGTTCGCTGCCGATAAACTGCCG 180
QY      61 ValIIeglyserProaspGluarGaspaAlaAlaLeuaAlaHisasnGluginleaspa 80
QY      ||| ||| ::| ||| ||||| ||||| |||
Db      181 CCGGCCGATTCGGCTGATGGCCAAAGCTGACGTGACGTACACAAATGGCAGATCACTGCCG 240
QY      81 LeuValgluthrArgAlaasnArgLeuTyrSergluginlglutThrProAlathrllea 100
QY      ||:::||||| |||||:::|||||::| ||||| ||||| |||||
Db      241 CTGATCGAGCGCGCCAGCCGCTGCACTTCGAAGGGGAAAACCCGGCAACCATCGCC 300

```


QY	101	GLuThrPheAlaLysAlaGluLysPheAspArgLeuAlaThrThrAlaSerSerAlaPhe	120
Db	301	GACACCTTCGCCAAGCGGAAAAAGCTGCAGCGATGGCGACGACTACATCAGCGCGTTCG	360
QY	121	GLuAsnThrProPheAlaAlaAlaSerValLeuGlnTyrMetGlnProAlaIleAsnLys	140
Db	361	CGGGCGACGCGCCCTTGCCATGGCCCTCGTTGCTTCAGTACATGCAGCCCTGCATCAACAAG	420
QY	141	GLyAspTrpLeuAlaThrProLeuLysProLeuThrProLeuIleSerGlyAlaLeuSer	160
Db	421	GGCGATTGGCTGCCGCGCTCCGCTCAAAACCGCTGACCCCGCTCATTTCCGGAGCGCTGTCCG	480
QY	161	GLyAlaMetAspGlnValGlyThrLysMetMetAspArgAlaArgGlyAspLeuHisTyr	180
Db	481	GGCGCCATGGACCAAGGTGGGCACCAAGATGATGACCGCGCAGCGGTGATCTGCATTAC	540
QY	181	LeuSerThrSerProAspLysLeuHisAspAlaMetAlaValSerValLysArgHisSer	200
Db	541	CTGAGCGCCTCGCCGACAGAGCTCCACGATGCAGATGGCCGCTTCGGTGAAGCGCCACTCG	600
QY	201	ProAlaLeuGlyArgGlnValValAspMetGlyIleAlaValGlnThrPheSerAlaLeu	220
Db	601	CCAAGCCTTGCTCGACAGGTTCTGGACACGGGGGTTGCGGTTCAAGACGTACTCGCGCCGC	660
QY	221	AsnValValArgThrValLeuAlaProAlaLeuAlaSerArgProSerValGlnGlyAla	240
Db	661	AACGCCGTACGTACCGTATTGGCTCCGGCACTGGCTCCAGACCCGCGTGACGGGTGCT	720
QY	241	ValAspPheGlyValSerThrAlaGlyGlyLeuValAlaAsnAlaGlyPheGlyAspArg	260
Db	721	GTGGACCTTGGTGATTCGATGCGGGGTGTGCTGGCTGCCAACGACGACTTTGGCAACGC	780
QY	261	MetLeuSerValGlnSerArgAspGlnLeuArgGlyGlyAlaPheValLeuGlyMetLys	280
Db	781	CTGCTCAGTGTGCAGTCGCGCTGATCACCAGCGTGGCGGTGCATTAGTCTCGGTTTGAAG	840
QY	281	AspLysGluProLysAlaAlaLeuSerGluGluThrAspTrpLeuAspAlaTyrLysAla	300
Db	841	GATAAAGAGCCCAAGGCTCAACTGAGCGAAGAAACGACTGGCTCGAGGCTTATAAGCA	900
QY	301	IleLysSerAlaSerTyrSerGlyAlaAlaLeuAsnAlaGlyLysArgMetAlaGlyLeu	320
Db	901	ATCAATCGGCCAGCTACTCGGGTCGGCGCCTCAACGCTGGCAAGCGATGCCGCTCTG	960
QY	321	ProLeuAspValAlaThrAspGlyLeuLysAlaValArgSerLeuValSerAlaThrSer	340
Db	961	CCACTGGATATGGCAGCCGACGCAATGGTGGCGGTAGAAGCCTGTGTACGCGTCCAGC	1020
QY	341	LeuThrLysAsnGlyLeuAlaLeuAlaGlyGlyTyrAlaGlyValSerLysLeuGlnLys	360
Db	1021	CTGACCCAAAACGGTCTGGCCCTGGCGGTGGCTTGCAGGGGTAGGCAAGTTGCAGGAG	1080
QY	361	MetAlaThrLysAsnIleThrAspSerAlaThrLysAlaAlaValSerGlnLeuSerAsn	380
Db	1081	ATGGCGACGAAAAATATCACCGACCCGGCGACCAAGCCCGGTCAGTCACTGACCAAC	1140
QY	381	LeuValGlySerValGlyValPheAlaGlyTyrThrThrAlaGlyLeuAlaThrAspPro	400
Db	1141	CTGGCAGGTTCCGCAAGCCGTTTTCGACAGGCTGACCAAGCGCGCTGACAAACGATCCC	1200
QY	401	AlaValLysLysAlaGlnSerPheIleGlnAspLysValLysSerThrAlaSerSerThr	420
Db	1201	GCGGTGAAAAAGCCGAGTCTTCATACAGAGACACGGTGAATCGACTGCATCCAGTACC	1260
QY	421	ThrSerTyrValAlaAspGlnThrValLysLeuAlaLysThrValLysAspMetSerGly	440
Db	1261	ACAGGCTACGTAGCCGACCAAGACCGTCAAACTGGCGAAGACCGTCAAAAGACATGGCGCGG	1320
QY	441	GLuAlaIleSerSerThrGlyAlaSerLeuArgSerThrValAsnAsnLeuArgHisArg	460
Db	1321	GAGGCGATCACCCATACCGGCGCCAGCTTGCGCAATACGGTCAATAACCTGCGTCAACGC	1380

```

OY      461  SETAlaProGluAlaAspIleGluGluGlyIleSerAlaPheSerArgSerGluThr 480
        |||  |||||||||||||||||||  ::|||  |||  |||||||
Db      1381  CCGGCTCGTGAAGCTGATATAGAGAGGGGCGACGGCGGCT--TCTCCAAGTGAATA 1437
OY      481  PropheGlnLeuArgArg 486
        ||||||::|  |||
Db      1438  CCGTTTCGGCCTATGCGG 1455

RESULT 4
AAD20405/c
ID      AAD20405 standard; DNA; 30365 BP.
XX
AC      AAD20405;
XX
DT      03-JAN-2002 (first entry)
XX
DE      P. syringae pv. tomato (Pto) DC3000 Conserved Effector Loci (CEL) DNA.
XX
KW      Conserved Effector Loci; CEL; cytostatic; antibacterial; gene therapy;
KW      Exchangeable Effector Loci; EEL; disease resistance; transgenic plant;
KW      eukaryotic cell death; cancer; ds.
XX
OS      Pseudomonas syringae.
PN      WO200175066-A2.
PD      11-OCT-2001.
PF      03-APR-2001; 2001WO-US10698.
XX
PR      03-APR-2000; 2000US-194160P.
PR      11-AUG-2000; 2000US-224604P.
PR      17-NOV-2000; 2000US-249548P.
XX
PA      (CORR ) CORNELL RES FOUND INC.
PA      (UYNE-) UNIV NEVADA & COMMUNITY COLLEGE SYSTEM.
PA      (UYNE-) UNIV NEBRASKA.
XX
PI      Collmer A, Alfano JR, Charkowski AO;
XX
DR      WPI; 2001-639361/73.
XX
PT      New nucleic acid molecules encoding proteins or polypeptides of
PT      Pseudomonas Conserved Effector Loci and Exchangeable Effector Loci
PT      genomic sequences, for imparting disease resistance to plants -
XX
PS      Disclosure; Page 9-17; 217pp; English.
XX
XX      The invention relates to an isolated nucleic acid molecule comprising a
CC      nucleotide sequence encoding proteins or polypeptides of Pseudomonas
CC      Conserved Effector Loci (CEL) and Exchangeable Effector Loci (EEL)
CC      genomic sequences. CEL and EEL DNA are useful for imparting disease
CC      resistance to a plant, by transforming a plant cell with the nucleic acid
CC      and regenerating a transgenic plant from the transformed plant cell,
CC      where the transgenic plant expresses a heterologous DNA molecule under
CC      conditions effective to impart disease resistance, or by treating a plant
CC      with an isolated protein or polypeptide, by applying the protein or
CC      polypeptide in an isolated form or by applying a non-pathogenic bacteria
CC      which secretes the protein or polypeptide, under conditions effective to
CC      impart disease resistance to the treated plant. CEL and EEL proteins
CC      are useful for causing eukaryotic cell death, by introducing a cytotoxic
CC      pseudomonas protein into a eukaryotic cell under conditions effective to
CC      cause cell death. CEL and EEL proteins are also useful for treating a
CC      cancerous condition, by introducing a cytotoxic Pseudomonas protein into
CC      cancer cells of a patient under conditions effective to cause death of
CC      cancer cells, and thus treating the cancerous condition. The method
CC      further involves administering a targetted DNA delivery system
CC      comprising a DNA molecule encoding the cytotoxic Pseudomonas protein,
CC      to the patient, where the targetted DNA delivery system delivers the
CC      DNA molecule into cancer cells and the cytotoxic Pseudomonas protein
CC      is expressed in the cancer cells. The present sequence is
XX      Pseudomonas syringae pv. tomato (Pto) DC3000 CEL DNA.
XX

```

SQ Sequence 30365 BP; 6828 A; 8938 C; 8516 G; 6080 T; 3 other;

Alignment Scores:

Pred. No.:	4,57e-146	Length:	30365
Score:	1869.50	Matches:	381
Percent Similarity:	85.19%	Conservative:	33
Best Local Similarity:	78.40%	Mismatches:	71
Query Match:	77.67%	Indels:	1
DB:	22	Gaps:	1

US-09-825-414-66 (1-487) x AAD20405 (1-30365)

QY 1 MethisIleasnGlnSerAlaGlnGlnProProGlyValAlaMetGlnSerPheArgThr 20
 Db 23984 ATGCACATCAACCGACGCGCTCCACACACCGCCTTGACTGCGACGCGATAGCTTTCGGACA 23925

QY 21 AlaserAspAlaSerIleuAlaSerSerSerValArgSerValSerThrThrSercysArg 40
 Db 23924 GCGTCCGACGCGCTCTTGCTGCCCTCCAGCTCTGTGCGATCTGTACAGCTCCGATCAGCAACGC 23865

QY 41 AspleuGlnAlaIleThrAspTyrIleuLysHisHisValPheAlaAlaHisArgPheSer 60
 Db 23864 GAGATAAATGCGATGCCGATTACCTGACAGATCATGTGTTGCGTCGCATAAACTGCCG 23805

QY 61 ValIleGlySerProAspGluArgAspAlaAlaIleuAlaHisasnGlnGlnIleAspAla 80
 Db 23804 CCGGCCGATTCGCTGATGGCCAAAGCTGCAGTTGACGTACACAAATGGCGCAGATCAGTCCG 23745

QY 81 LeuValGluThrArgAlaAsnArgIleuTyrSerGluGlyGluThrProAlaThrIleAla 100
 Db 23744 CTGATCGAGAGCGGCGCCAGCGCCTGCACCTTCGAAGGGAAACCCGGCAACCATCGCC 23685

QY 101 GluThrPheAlaLysAlaGluLysPheAspArgLeuAlaThrThrAlaSerSerAlaPhe 120
 Db 23684 GACACCTTCGCCAAGCGGAAAGCTGCACCGATTGGCGACGACTACATCAGCGCGCTTG 23625

QY 121 GluAsnThrProPheAlaAlaAlaSerValLeuGlnTyrMetGlnProAlaIleAsnLys 140
 Db 23624 CGGGCGACGCCCTTTGCCATGGCCCTCGTTGCTTCAGTACATGACGCTGCGATCAACAAG 23565

QY 141 GlyAspTyrPleuAlaThrProLeuLysProLeuThrProLeuIleSerGlyAlaLeuSer 160
 Db 23564 GCGGATTTGGCTGCCGCGCTCCGCTCAACCGCTGACCCCGCTCATTTCCGGAAGCGCTGCG 23505

QY 161 GlyAlaMetAspGlnValGlyThrLysMetMetAspArgAlaArgGlyAspLeuHisTyr 180
 Db 23504 GCGCGCATGGACCAAGGTGGGCACCAAGATGATGACCGCGCAGCGGTGATCTGCATTAC 23445

QY 181 IeuSerThrSerProAspLysLeuHisAspAlaMetAlaValSerValLysArgHisSer 200
 Db 23444 CTGAGCGCTGCGCGGACACAGGCTCCACGATGGCGCTTCGGTGAAGCGCCACTCG 23385

QY 201 ProAlaLeuGlyArgGlnValValaAspMetGlyIleAlaValGlnThrPheSerAlaLeu 220
 Db 23384 CCAAGCCTTGCCTCGACAGAGTTCTGACACAGGGGGTTGCGGTTCAAGACTACTCGGGCGCG 23325

QY 221 AsnValValArgThrValLeuAlaProAlaLeuAlaSerArgProSerValGlnGlyAla 240
 Db 23324 AACGCCGTACGTACCGCTATTGGCTCCGGCACTGGCGCTCCAGACCCCGCTGCAGGGTGCT 23265

QY 241 ValAspPheGlyValSerThrAlaGlyGlyLeuValAlaAsnAlaGlyPheGlyAspArg 260
 Db 23264 GTGACCTTGGTGTATCGATGGCGGGTGTGCTGCCAACGCAAGGCTTTGGCAACCGC 23205

QY 261 MetLeuSerValGlnSerArgAspGlnLeuArgGlyGlyAlaPheValLeuGlyMetLys 280
 Db 23204 CTGCTCAGTGTGAGTGCAGTGCATCACCAGCGTGCGGTGATAGTGTGCGGTTTGAAG 23145

QY 281 AspLysGluProLysAlaAlaLeuSerGluGluThrAspTyrPleuAspAlaTyrLysAla 300
 Db 23144 GATAAAGAGCCCAAGGCTCAACTGAGCGAAGAAACGACTGCGTCGAGGCTTATAAAGCA 23085

QY 301 IleLysSerAlaSerTyrSerGlyAlaAlaLeuAsnAlaGlyLysArgMetAlaGlyLeu 320

Db 23084 ATCAAAATCGCCACGACTACTCGGGTCCGGCGCTCAACGCTGGCAAGCGGATGGCGGTTG 23025

QY 321 ProIeuAspValAlaThrAspGlyLeuLysAlaValArgSerIleuValSerAlaThrSer 340
 Db 23024 CCACGTGATATGGCGGACCGACGCAATGGGTGCGGTAAAGAACCTGGTGTACGCTCCAGC 22965

QY 341 LeuThrLysAsnGlyLeuAlaLeuAlaGlyGlyTyrAlaGlyValSerLysLeuGlnLys 360
 Db 22964 CTGACCCCAAAACGGTCTGGCGCCCTGGCGGGTGGCTTTCAGGGGTAGGCAAGTTCAGAG 22905

QY 361 MetAlaThrLysAsnIleThrAspSerAlaThrLysAlaAlaValSerGlnLeuSerAsn 380
 Db 22904 ATGGCGAGCAAAATATACACCGACCGCGGCAAGCGCGGCTCAGTCACTGACCAAC 22845

QY 381 LeuValGlySerValGlyValPheAlaGlyTyrThrThrAlaGlyLeuAlaThrAspPro 400
 Db 22844 CTGGCAGGTTCCGACAGCCGTTTTCGACAGGCTGGACACGCGCGCTGACAAACGATCC 22785

QY 401 AlaValLysLysAlaGluSerPheIleGlnAspLysValLysSerThrAlaSerSerThr 420
 Db 22784 GCGGTGAATAAAGCCGAGTCTGTCTATACAGACAGCAGGTGAATCGACTGCATCCAGTACC 22725

QY 421 ThrSerTyrValAlaAspGlnThrValLysLeuAlaLysThrValLysAspMetSerGly 440
 Db 22724 ACAGGCTACGTAGCCGACCAAGACCGGTCAAACTGGCGAAGACCGGTCAAAACATGGCGCG 22665

QY 441 GluAlaIleSerSerThrGlyAlaSerLeuArgSerThrValAsnAsnLeuArgHisArg 460
 Db 22664 GAGGCGATCACCATACCGGCGCCAGCTTCCGAATACGGTCAATACCTGCGTCAACGC 22605

QY 461 SerAlaProGluAlaAspIleGluGlyGlyIleSerAlaPheSerArgSerGluThr 480
 Db 22604 CCGGCTCGTGAAGCTGATATAGAAGGGGGGACAGCGGCGT---TCTCAAGTGAATA 22548

QY 481 ProPheGlnLeuArgArg 486
 Db 22547 CCGTTTCGGCCTATGCGG 22530

RESULT 5
 ABL29757
 ID ABL29757 standard; DNA; 5397 BP.
 XX
 AC ABL29757;
 XX
 DT 26-MAR-2002 (first entry)
 XX
 DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 40744.
 XX
 KW Drosophila; developmental biology; cell signalling; insecticide;
 KW pharmaceutical; gene; ds.
 XX
 OS Drosophila melanogaster.
 XX
 PN WO200171042-A2.
 XX
 PD 27-SEP-2001.
 XX
 PF 23-MAR-2001; 2001WO-US09231.
 XX
 PR 23-MAR-2000; 2000US-191637P.
 PR 11-JUL-2000; 2000US-061415O.
 XX
 PA (PEKE) PE CORP NY.
 XX
 PI Venter JC, Adams M, Li PWD, Myers EW;
 XX
 DR WPI; 2001-656860/75.
 XX
 PT New isolated nucleic acid detection reagent for detecting 1000 or more
 PT genes from Drosophila and for elucidating cell signalling and cell-cell
 PT interactions -
 XX

PS Claim 1; SEQ ID NO 40744; 21pp + Sequence Listing; English.
XX The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABBS7737-ABBS72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX

SQ Sequence 5397 BP; 1698 A; 1618 C; 1178 G; 903 T; 0 other;

Alignment Scores:

Pred. No.:	0.019	Length:	5397
Score:	148.50	Matches:	102
Percent Similarity:	31.09%	Conservative:	64
Best Local Similarity:	19.10%	Mismatches:	231
Query Match:	6.17%	Indels:	137
DB:	23	Gaps:	17

US-09-825-414-66 (1-487) x ABL29757 (1-5397)

QY 12 GlyValAlaMetGluSerPheArgThrAlaSerAspAlaSerLeuAlaSerSerVal 31
DB 3 GSTACAGCAGCAGCAGATAGTTCAGCAACAGCAAAATTGTTACAGCAGCAAGTTGTCAGCA 62
QY 32 ArgSerValSerThrThrSerCysArgAspLeuGlnAlaIleThrAspTyrLeuLysHis 51
DB 63 ACAGCAGCCACACACAACTCTGTCCGGT----- 92
QY 52 HisValPheAlaAlaHisArgPheSerValIleGlySerProAspGluArgAspAlaAla 71
DB 93 -----TTGCGCCGAAATCAT-----ACACCAACATCTAATGCGGCAGCA 131
QY 72 LeuAlaHisAsnGluGlnIleAspAlaLeuValGluThrArgAlaAsnArgLeuTyrSer 91
DB 132 ACAAAATTTGAACCAAGATCAATGCAATGTCAGCAAAATGTCGCAAGCGGACACCATT 191
QY 92 GluGluGluThrProAla-----ThrIleAlaGluThrPheAla 104
DB 192 GACGCCGCAACAGCAGCAGCAACAGAGACAATTCACACACAGCTGCACACAAAGTTATGCA 251
QY 105 LysAlaGluLysPheAspArgLeuAlaThrThrAlaSerSerAlaPheGluAsnThr--- 123
DB 252 GCAGCAGCAGCAATTACAGATGCAAAATGCAGCAGCAGCAACAGCAACAGAAATCTACACA 311
QY 124 -----ProPheAlaAlaAlaSerValLeuGlnTyrMet 134
DB 312 ACAACAGTCGCTCGATGGTGCAAAACCGTTC--GCCGGTAATGCCACCAAGTACGCC 368
QY 135 GlnProAlaIleAsnLysGly-----AspTyrLeuAlaThrProLeu 148
DB 369 ATCCCATCGTTGCAGCAGCAACAGCATCATGCAACTAGCAATATATGTTGCCGCAACATC 428
QY 149 LysProLeuThrProLeuIleSerGlyAlaLeuSerGlyAlaMetAspGlnValGlyThr 168
DB 429 GCCCGCAGCACTGCAATCCCCAGCTCCCAATAGTCCGCCACCTCCGCCCTCGCGCCA 488
QY 169 LysMetMetAspArgAlaArgGlyAspLeuHisTyrLeuSerThrSerProAspLysLeu 188
DB 489 AAACGCCCAGCAACACATTTGGCCAGCAACAGCAAAATGCAACAGAAATGCACAAGCA 548
QY 189 HisAspAlaMetAlaValSerValLysArgHisSerProAlaLeuGlyArgGlnValVal 208
DB 549 CATGATGGGCTCGCCACACAACC-----TCAACCAACATCTAT 587
QY 209 AspMetGlyIleAlaValGlnThrPheSerAlaLeuAsnValValArg----- 224
DB 588 GATGTGCGCCGCAACAGCAGCAACAGATGCAAGCCCTTTCACAGCCCGTCCATCAACAGCA 647

QY 225 -----ThrValLeuAlaProAlaLeuAlaSerArgProSerVal 237
DB 648 GCGCATGCAATTGCAGCAGCAACAACAATTTGGCCCAACAGCAGCAACAACAGCAGCA 707
QY 238 GlnGly-----AlaValAspPheGlyValSerThrAlaGlyGly 250
DB 708 GAGTCCCCAGCATATTTCTCCGCAATCGCCCAAGATTCTCAAACTCCACCACATGCAGGC 767
QY 251 LeuValAlaAsnAlaGlyPheGlyAspArgMetLeuSerValGlnSerArgAspGlnLeu 270
DB 768 AAAGCTGCATCAGCATCAGGTGGT-----TCCTGACAAAGCCAC 806
QY 271 ArgGlyGlyAlaPheValLeuGlyMetLysAspLysGluProLysAla----- 286
DB 807 TCAGCCCGCAGCAATCATCTCTCCAGCAGAAAGCCTATCGACCCACAGGCCCGTTCAAGT 866
QY 287 -----AlaLeuSerGluGluThrAspTyrPheAspAlaTyrLys 299
DB 867 AGCTCAAGTGTCTCAGTCGATCTGCATTTGAGCAGCAATCAGGACTCGTTGATTAAT----- 920
QY 300 AlaIleLysSerAlaSerTyrSerGlyAlaAlaLeuAsnAlaGlyLysArgMetAlaGly 319
DB 921 -----GAGCAGCAACAATTGAAGCAGCAACAACAGATGCCAACAGCAACAATGGC 974
QY 320 LeuProLeuAspValAlaThrAspGlyLeuLysAlaValArgSerLeuValSerAlaThr 339
DB 975 ACCGCAAGCCCGCAGCAGCAACAATGGCACAGCAACCGCAACAGCAGCAGCAACCA 1034
QY 340 SerLeuThrLysAsn-----GlyLeuAlaLeuAlaGlyGlyTyr--- 352
DB 1035 ACAACAGCAGCAGCACCCCTTCTCCGCGACAATCGCCACTGCAACAGCAACCTACTACCCC 1094
QY 353 -----AlaGlyValSerLysLeuGlnLysMetAlaThrLysAsnIleThrAspSerAla 370
DB 1095 CACCTTGCACAGCAGCCCAAAACCAAAACCGCTTCACAAATACAGCAGCAACACACA 1154
QY 371 ThrLysAlaAlaValSerGlnLeuSerAsnLeuValGlySer-----Val 385
DB 1155 ACAACAGCAGCAGCAGCAGCAACAACAGCAGCAGCAACAACAGCAGGTCTGACGCA 1214
QY 386 GlyValPheAlaGlyTyrThrAla----- 394
DB 1215 GCAACAGCCACAGCCTGGACAACAACAAGTCAATGCTCAGCGCATGTTATAACAC 1274
QY 395 -----GlyLeuAlaThrAsp 399
DB 1275 ATCCACGGCTCAAGGACAACAATCATTCAAAGTCAATGCTTTGGCCTTGCAAGACA 1334
QY 400 ProAlaValLysLysAlaGluSerPheIleGlnAspLysValLysSerThrAlaSerSer 419
DB 1335 GCAGCAAT-----GTTGCACGTTCAACAACAACAGCA 1367
QY 420 ThrThrSerTyrValAlaAspGlnThrValLysLeuAlaLysThrValLysAspMetSer 439
DB 1368 ACAGCAGCCGCGCAGCAGCAACAACAGATTACTCAAGTGCAGCAACTGCCACCGGCAGCA 1427
QY 440 GlyGluAlaIleSerSerThrGlyAlaSerLeuArgSerThr 453
DB 1428 GCAGCAACAGCTTCGCAACAGCAACCAAGTTCAACAACAACA 1469
RESULT 6
AAD31881
ID AAD31881 standard; DNA; 4645 BP.
XX
AC AAD31881;
XX
DT 18-JUN-2002 (first entry)
XX
DE Lactobacillus rhamnosus outer membrane protein rompa' gene.
KW Enzyme; flavour; aroma; texture; nutritional; dairy manufacture; therapy;
KW fermentation process; anti-infection; rotavirus infection; heart disease;

KW infantile diarrhoea; lactose digestion; anti-cancer; autoimmune disorder;
KW anti-mutagenesis; immune system modulation; allergy; Helicobacter pylori;
KW antihypertensive effect; urogenital infection; hepatic encephalopathy;
KW bowel syndrome; endocarditis; transgenic microbe; outer membrane protein;
KW rompa gene; ds.
XX
OS Lactobacillus rhamnosus HN001.

XX Key Location/Qualifiers
FH CDS 74..4465
FT /*tag= a
FT /product= "Outer membrane protein"
FT /note= "CDS does not include start codon"
FT /partial

XX PN W0200212506-A1.
XX PD 14-FEB-2002.
XX PF 08-AUG-2001; 2001WO-NZ00160.
XX PR 08-AUG-2000; 2000US-0634238.
XX PR 28-NOV-2000; 2000US-0724623.
XX PA (GENE-) GENESIS RES & DEV CORP LTD.
XX PA (VIAL-) VIALACTIA BIOSCIENCE NZ LTD.
XX PI Glenn M, Havukkala J, Bloksberg LN, Lubbers MW, Dekker J;
XX PI Christensson AC, Holland R, O'Loole PW, Reid JR, Coolbear T;
XX WPI: 2002-241760/29.
XX DR P-PSDB; AAE20110.

XX PT New polynucleotides and polypeptides from Lactobacillus rhamnosus,
PT useful in e.g. improving the flavor, aroma, texture and health-related
PT benefits of milk-derived products, or in increasing properties of
PT microbes -
XX Claim 2; Fig 63; 257pp; English.

XX CC The present invention relates to a new isolated polynucleotide comprising
CC a sequence present in Lactobacillus rhamnosus strain HN001 and encoding a
CC polypeptide capable of modifying the flavour, aroma, texture, nutritional
CC and health benefits of milk-derived products, and/or survivability of
CC microbes in dairy manufacturing processes. The polynucleotides are useful
CC for improving the properties of microbes used in the manufacture of milk-
CC derived products such as cheeses, yogurt, fermented milk products, sour
CC milks and buttermilk; in modifying the flavour, aroma, texture and health
CC -related benefits of milk-derived products and in increasing the survival
CC of microbes during industrial fermentation processes. The bacteria may be
CC used to increase resistance to enteric pathogens and anti-infection
CC activity, including treatment of rotavirus infection and infantile
CC diarrhoea; aid in lactose digestion; as anti-cancer and anti-mutagenesis;
CC liver cancer reduction; reduction of small bowel bacterial overgrowth;
CC immune system modulation and treatment of autoimmune disorders and
CC allergies; treatment of allergic responses to foods; reduction of blood
CC lipids and prevention of heart disease; antihypertensive effect;
CC prevention and treatment of urogenital infections, Helicobacter pylori,
CC or hepatic encephalopathy; treatment of inflammatory bowel disorder and
CC irritable bowel syndrome; modulation of endocarditis; and for improved
CC protein and carbohydrate utilization and conversion. The transgenic
CC microbial population can be administered to a mammal as an anti-
CC carcinogenic agent. The present sequence is Lactobacillus rhamnosus
CC outer membrane protein rompa gene.

XX SQ Sequence 4645 BP; 1279 A; 1062 C; 1261 G; 1043 T; 0 other;

Alignment Scores:

Pred. No.: 0.021 Length: 4645
Score: 147.00 Matches: 104
Percent Similarity: 33.53% Conservative: 64
Best Local Similarity: 20.76% Mismatches: 241
Query Match: 6.11% Indels: 92

DB:	24	Gaps:	12
US-09-825-414-66 (1-487) x AAD31881 (1-4645)			
QY 1 MethiSlleAaInSerAlaGlnInProProGlyValAlaMetGlSerPheArgThr 20			
Db 2213 ATCAACGCCAATAGTGTCTGCTGATGTGCACAGCTTCACAGGCAAGCCTACAGTGTGCT 2272			
QY 21 AlasSerAspAlaSerLeuAlaSerSerSerValArgSerValSerThrThrSerCysArg 40			
Db 2273 GCTAGTGAATGGCGGAGTATGCAAGTGAAGCCCAATCGATTCGTGCAGTCATGCTGAT 2332			
QY 41 AspleuGlnAlaIleThrAspTyrLeuYshHisValPheAlaAlaHisArgPheSer 60			
Db 2333 AATATGGAA-----ATCAAGTCTCTCGCCAGTGAATGCTGAGAAGCAATCG 2377			
QY 61 ValIleGlySerProAspGluArgAspAlaAlaLeuAlaHisAsnGlnIleAspAla 80			
Db 2378 CAAATTGCTTTGGCAGCTAGCAAGTCTGCTGCGCTAGTTCACGTGCGGCAAGCTCCGCA 2437			
QY 81 LeuValGluThrArgAlaAsnArgLeuTyrSerGluGlyGluThrProAlaThrIle--- 99			
Db 2438 GCAATCGTG-----GCAAGTAGCGGCGCTAGTGAAGCGTCATCTGCAGCTGCGCGTA 2491			
QY 100 -----AlaGluThrPheAlaAlaGluLysPheAspArgLeuAlaThrThr 115			
Db 2492 AGTAACGCTGATGCATCAGCAAACTCTGCAGCGCTGCTATGATTCCTACGCTTCTGAG 2551			
QY 116 AlasSerSerAla-----PheGluAsnThrProPheAlaAla 127			
Db 2552 GCCAGTCCCGCTTCTGCTGCTAATGATGATTCGGGATATGCCACTGCATCATTTGCAGCA 2611			
QY 128 AlasSerValLeuGlnTyrMetGlnProAlaIleAsnLysGlyAspTyrPleuAlaThrPro 147			
Db 2612 AGTTCCCGCTGCGCTGCCATG----- 2632			
QY 148 LeuLysProLeuThrProLeuIleSerGlyAlaLeuSerGlyAlaMetAspGlnValGly 167			
Db 2633 -----AGCGCAGCGTTATCGACAGCGCAAGTTGCTGCCAAG 2668			
QY 168 ThrLysMetMetAspArgAlaArgGlyAspLeuHisTyrLeuSerThrSerProAspLys 187			
Db 2669 GTTGCACTGATGATGCAGCAGCAGCGGCT----- 2698			
QY 188 LeuHisAspAlaMetAlaValSerValLysArgHisSerProAlaLeuGlyArgGlnVal 207			
Db 2699 -----AGTGCAGCTGCTGTTGCTAGTGCAGCTCAAGCGACTCCAGAATTAACAAGCG 2752			
QY 208 ValAspMetGlyIleAlaValGlnThrPheSerAlaLeuAsnValValArg-----Thr 225			
Db 2753 ACTGCAGCTACAGCAAGAAGTCAAGCAGCTGATGATTTGAATAGATCAAGTCTTAAGT 2812			
QY 226 ValLeuAlaProAlaLeuAlaSerArgProSerValGlnGly----- 239			
Db 2813 GATTACGCAAGTGGCGCAAGCTCCAGTCCAGCGGAAGCGGTCAAGCATCGACTGCACACA 2872			
QY 240 -----AlaValAspPheGlyValSerThrAlaGlyGlyLeuValAlaAsn 254			
Db 2873 TCTGCGTATGCTAGTGTGTCAGAGTTCAGTCCAGTGAAGCCGGTTCATATGCTCATCAG 2932			
QY 255 AlaGlyPheGlyAspArgMetLeuSerValGlnSerArgAspGlnLeuArgGlyGlyAla 274			
Db 2933 GCAGGCTCAGCGCCAGTGACGCTGTGCGTCACTCCGCGAGTGCAGCCCAACATGCCAGC 2992			
QY 275 PheValLeuGlyMetLysAspLysGluProLysAlaAlaLeuSerGluGluThrAspTyr 294			
Db 2993 ACCGCTGCGAGTGGCGCATCCAGCTATCCGAAGATAGTGGATTCAGTCACTAGCCAGT 3052			
QY 295 LeuAspAlaTyrLysAlaIleLysSerSerAlaSerTyrSerGlyAlaAlaLeuAsnAlaGly 314			
Db 3053 CAGGCTGCAAGCAGGACGCAAAAGGCAAGCAGTAACGCGAGTCCCGCAACAGCGCC--- 3109			
QY 315 LysArgMetAlaGlyLeuProLeuAspValAlaThrAspGlyLeuLysAlaValArgSer 334			

Db 3110 -----GCGGCGCTGGTTTCAGTGCCTGCCAGTGCATGCACAGTGAACAGGCGGAAGACG 3160
QY 335 LeuValSerAlaThrSerLeuThrLysAsnGlyLeuAlaLeuAlaGlyGlyTyrAlaGly 354
Db 3161 GCTGCAAGTCCGATGTGGTGCAAGCAGTGGCGCCAGCAGCGCTAACAGTAATCGGAGT 3220
QY 355 ValSerLysLeuGlnLysMetAlaThrLysAsnIleThrAspSerAlaThrLysAlaAla 374
Db 3221 GCCGCAGCC-----AGTGCACCAAGCGTGGT 3247
QY 375 ValSerGlnLeuSerAsnLeuValGlySerValGlyValPheAlaGlyTyrPThrAla 394
Db 3248 GATAGCAAAAGCCGCA-----GCAGCATTCCTCGAGTGCA 3280
QY 395 GlyLeuAlaThrAspProAlaValLysLysAlaGlnSerPheIleGlnAspLysValLys 414
Db 3281 GCGAGTGTGTCAGCAGCAGTGGCCAGGGGTGCAGACAGTTCGCCAGCAGCGCGGAGT 3340
QY 415 SerThrAlaSer-----SerThrThrSerTyrValAlaAspGlnThrValLysLeuAla 432
Db 3341 GCCGCGGCATCCGATGACTCGGTAGCTTCTAGTCCGCCAGTGGCGCTGCAGGCTTTGAC 3400
QY 433 LysThrValLysAspMetSerGlyGlyAlaIleSerSerThrGlyAlaSerLeuArgSer 452
Db 3401 AAAGCTGCCAGCGCTGCCGAAGGCGCAGCTTCAGTCCCGCAGCGCGGCTCTAGTTCA 3460
QY 453 ThrValAsnAsnLeuArgHisArgSerAlaProGluAlaAspIleGluGlyGlyIle 472
Db 3461 GCGGCAGCTCAAGGCACACAGAGTGGCGCAAGCTCCAGTGCCAGCAGCGGCTCAAGCA 3520
QY 473 Ser 473
Db 3521 TCA 3523

RESULT 7
ABL23378/c
ID ABL23378 standard; DNA; 3146 BP.
XX
XX ABL23378;
AC
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster genomic polynucleotide SEQ ID NO 21607.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical; gene; ds.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Claim 1; SEQ ID NO 21607; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is

CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins
CC (ABB57737-ABB72072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct-sequences.
XX

SQ Sequence 3146 BP; 783 A; 811 C; 727 G; 825 T; 0 other;

Alignment Scores:

Pred. No.:	0.0141	Length:	3146
Score:	146.50	Matches:	91
Percent Similarity:	39.32%	Conservative:	60
Best Local Similarity:	23.70%	Mismatches:	156
Query Match:	6.09%	Indels:	77
DB:	23	Gaps:	13

US-09-825-414-66 (1-487) x ABL23378 (1-3146)

QY 141 GlyAspTrpLeuAlaThrProLeuLysProLeuThrProLeuIleSerGlyAlaLeuSer 160
Db 2255 GCGGCTTAGTGGCCAGCAGAGCTTCAGACGACGACGAGCTTGTAGCGGTAGCCATCGCT 2196
QY 161 GlyAlaMetAspGlnValGlyThrLysMetMetAspArgAlaArgGlyAspLeuHis--- 179
Db 2195 GGCAGGAGCGTCTGCTCGCGGCGCACATCATGATGGCAGAGCGGCTCTCTCCACGGG 2136
QY 180 -----TyrLeuSerThrSerProAspLysLeu 188
Db 2135 GGCAGGTACTCGCCGAAGCAGAGCTCAGAGACATCACGGCGCAGCATCACAGCAGC 2076
QY 189 HisAspAlaMetAlaValSerVal-----LysArgHisSerProAlaLeuGly 204
Db 2075 CACGGTCTTGTAGCGGTATCCATCATCAGCCAGGGAGTCTCCACAGCAACAGCCTCGGA 2016
QY 205 ArgGlnValValAspMetGlyIleAlaValGln-----ThrPheSer 218
Db 2015 GCGGACTCCTCAACAGGGGGCAGGTATTTCAGGGGACAGCTCGGAGACATCACAGCAGC 1956
QY 219 AlaLeuAsnValValArgThrValLeuAlaProAlaLeuAlaSerArgProSerValGln 238
Db 1955 ACGGATCACACGACGACGGCTTGTAGCGGTATCCATCATCAGCCAGAGAGTCTCGAC 1896
QY 239 GlyAlaValAspPheGlyValSerThrAlaGlyGlyLeuValAlaAsnAlaGlyPheGly 258
Db 1895 GGC-----AACGGCTTCGGA 1881
QY 259 AspArgMetLeuSerValGlnSerArgAspGlnLeuArgGlyGlyAlaPheValLeuGly 278
Db 1880 GCGGATTCCTC-----AACAGGGGCGAGTATTCGGCGGACAGCTCGTTAC--- 1833
QY 279 MetLysAspLysGluProLysAlaAlaLeuSerGluGluThrAspTriPleu----- 295
Db 1832 -----ATCACGACGGCGCAGCGATCACAGCAGGACAGTCTGTAGCGGTA 1788
QY 296 -----AspAlaTyrLysAlaIleLysSerAlaSerTyrSerGlyAlaAlaLeuAsnAla 313
Db 1787 TCCATCAGCAGCCAGAGGGGTCTCAACAGCAACACGCTCGGAGGACAGACTCTCTCAACGGG 1728
QY 314 GlyLysArgMetAlaGlyLeuProLeu---AspValAlaThrAspGlyLeuLysAlaVal 332
Db 1727 GGGCAGGTACTCGCCGAAGGAGGAGCTCATTCAGATTCACGACGACGAGTCAACGGCG 1668
QY 333 ArgSerLeuValSerAlaThrSerLeuThrLysAsnGlyLeu-----AlaLeuAla 349
Db 1667 GACGGTCTTGTAAACGTTAAACATCATCAGCAAGAGGGGTCTCCACAGCAACAGCCTCAGA 1608
QY 350 GlyGlyTyrAlaGlyValSerLysLeuGlnLysMetAlaThrLysAsnIleThrAspSer 369
Db 1607 GCGGAGTCTCTCAACAGCGGCGCAGGTATTTCAGCGGACAGCTCGAACAATCAGCAGCAGC 1548

QY 370 AlaThrLysAlaAlaValSerGlnLeuSerAsnLeuValGlySerValGly---ValPhe 388
Db 1547 ACGGATCACACGCGGACAGTCTTGTAGCGGTAGCCATCATCAGCCAGGGAGTCTCGAC 1488
QY 389 AlaGlyTrpThrThrAlaGlyLeuAlaThrAspProAlaValLysLysAlaGluSerPhe 408
Db 1487 AGCAACAGCCTCGGAGCGGACTCTCAACAGGGGGCAGGTACTCAGGGAGAGTTTC--- 1431
QY 409 IleGlnAspLysValLysSerThrAlaSerSerThrThrSerThrValAlaAspGlnThr 428
Db 1430 -----GGAGACATCACGACGACGACGATCATCAGCGGAGGAGTTC 1389
QY 429 ValLysLeuAla-LysThrValLysAspMetSer-----GlyGluAlaIleSe 444
Db 1388 GTAACGGTAGCCATCATCAGCCAGAGAGTCTCGACGCGCAACGGCTTCGAGCGGATTC 1329
QY 444 rSerThr-GlyAlaSerLeuArgSerThrValAsnLeuArgHisArgSerAlaProG 464
Db 1328 CTCACACAGGGGGCAGGTAT-----CACCGAGGGGAGCTC 1293
QY 464 IuAlaAspIleGluGluGlyGlyIleSerAlaPheSerArgSerGluThrProPheGlnL 484
Db 1292 GTTCACATCACGACGCGTGGCGCACTTCAGCGCGCGCACGGTCTGTAGCGGTATCCATC 1233
QY 484 euArgArg 486
Db 1232 GTCGGAGA 1225
RESULT 8
AAD04029
ID AAD04029 standard; DNA; 9542 BP.
XX AAD04029;
AC AAD04029;
XX 02-JUL-2001 (first entry)
DT
XX Moraxella catarrhalis outer membrane protein-106 (OMP106) DNA.
DE
XX Outer membrane protein-106; OMP106; haemagglutination; vaccine;
KW bacterial infection; immunogen; cytotoxic; antibiotic;
KW passive immunisation; ds.
XX
XX Moraxella catarrhalis.
XX
FH key location/Qualifiers
FT CDS 218..6589
FT /*tag= a
FT /product= "Outer membrane protein-106 (OMP106)"
XX
PN US6214981-B1.
XX
PD 10-APR-2001.
XX
PF 12-NOV-1997; 97US-0968685.
XX
PR 03-MAY-1996; 96US-0642712.
XX
PA (ANTE-) ANTEX BIOLOGICS INC.
XX
PI Tucker K, Plosila L, Tillman UF;
XX
DR WPI; 2001-281002/29.
DR P-PSDB; AAE00701.
XX
XX Novel nucleotide sequences encoding Moraxella catarrhalis outer
PT membrane protein-106 polypeptide, useful for diagnosis of bacterial
PT infections and as vaccine against Moraxella catarrhalis infection of
PT mammals -
XX
PS Claim 6; Column 45-54; 49pp; English.
XX
CC The present sequence is haemagglutinating Moraxella catarrhalis outer

CC membrane protein-106 (OMP106) DNA. The OMP106 is used as a therapeutic
CC and prophylactic vaccine against M. catarrhalis infections of mammals.
CC It is used for diagnosis of bacterial infections and as reagents for
CC clinical or medical diagnosis of M. catarrhalis infections and for
CC scientific research on the properties of pathogenicity, virulence and
CC infectivity of M. catarrhalis. It is also used as a probe to identify
CC the presence of M. catarrhalis in biological specimens and to identify
CC other bacteria that encode a polypeptide related to M. catarrhalis
CC OMP106. OMP106-derived polypeptides are used as ligands to detect
CC antibodies elicited in response to M. catarrhalis infections and also
CC as immunogens for inducing M. catarrhalis-specific antibodies which are
CC useful in immunoassays to detect M. catarrhalis in biological specimens.
CC Cytotoxic antibodies are useful in passive immunisations against
CC M. catarrhalis.
XX
SQ Sequence 9542 BP; 2966 A; 2104 C; 2098 G; 2374 T; 0 other;
Alignment Scores:
Pred. No.: 0.0578 Length: 9542
Score: 146.50 Matches: 111
Percent Similarity: 36.00% Conservative: 69
Best Local Similarity: 22.20% Mismatches: 181
Query Match: 6.09% Indels: 139
DB: 22 Gaps: 23
US-09-825-414-66 (1-487) x AAD04029 (1-9542)
QY 32 ArgSerValSerThrThrSerCysArgAspLeuGlnAlaIleThrAspTyrLeuLysHis 51
Db 1346 AAGACTTTAACTATCATCAGGTGGTGCCACAGACCAGCGCATTAACCGAT----- 1393
QY 52 HisValPheAlaAlaHisArgPheSerValIleGlySerProAspGluArgAspAlaAla 71
Db 1394 -----CATACATCGGTGTGTGTACAAATGGCGATGGTCTGAAGTTCAA 1438
QY 72 LeuAlaHisAsnGluGlnIleAspAlaLeuValGluThrArg-----AlaAsnArg 88
Db 1439 CTGTGTAAGCTTTAACCAGCGCTTAAATGGTTACCACCTGAACCTAACCGCCAACGAG 1498
QY 89 LeuTyrSerGluGlyGluThrProAlaThrIleAlaGluThrPheAlaLysAlaGluLys 108
Db 1499 AAAGTTACCGTAGCAAAACC----- 1519
QY 109 PheAspArgLeuAlaThrThrAlaSerSerAlaPheGluAsnThrProPheAlaAlaAla 128
Db 1520 -----CGCCTT---ACCACAGATTAATAATGGTTTACCAATGAT-----ATG 1558
QY 129 SerValLeuGlnTyrMetGlnProAlaIleAsnLysGlyAspTyrPheAlaThrProLeu 148
Db 1559 AATGGCATTTGATGAAGCAAACTTATCTTGATAAAGACACTGGCATTCATGCAGGTGGT 1618
QY 149 LysProLeuThrProLeuIleSerGlyAlaLeu----- 159
Db 1619 CAAAAGATTACCAAACTTACTGTGTGTAGATGACGATCGGCAACTTATGACAG 1678
QY 160 -----SerGlyAlaMetaspGlnValGlyThrLysMetMet 171
Db 1679 CTTAAAAAGTTAACCACCAACCGCTGAAGTGTCTACAAACCTTTACCGTTAAAAAGTGA 1738
QY 172 AspArgAlaArgGlyAspLeuHisTyrLeuSerThrSerProAspLysLeuHisAspAla 191
Db 1739 GATAAAATGGTAATGAT-----GCTAATGACAGC 1768
QY 192 MetAlaValSerValLysArgHisSerProAlaLeuGlyArgGlnValValAspMet--- 210
Db 1769 AAAATCATCACCGTGGTAAATAATACAAACAGACGCTACTCAAGTCAACACCTTAATA 1828
QY 211 -----GlyIleAlaValGlnThrPheSerAlaLeuAsnValValArgThr 225
Db 1829 CTCAAAGGTGAACGCGTGTGATGTTACAACC----- 1861
QY 226 ValLeuAlaProAlaLeuAlaSerArgProSerValGlnGlyAlaValAspPheGlyVal 245

QY	142	-----AspTrpLeuAlaThrProLeuLysProLeuThrProLeuIleSer	156
Db	6365	GCATCATGCAACTAGCAATATGTTGGCCGCACAAATCGCCGACAACTGCAATCCACAGC	6306
QY	157	GLYAlaLeuSerGlyAlaMetaspGlnValGlyThrLysMetMetAspArgAlaArgGly	176
Db	6305	TCCCCAAATGACTCCGCCACCTCCGCCCTCGCCGCAAAAGCCCAAGCAACAACAATTGGCG	6246
QY	177	AspLeuHisTyrLeuSerThrSerProAspLysLeuHisAspAlaMetAlaValSerVal	196
Db	6245	CCAGCAACAGCAAAATGCAACAGATCGACACAGCACATGATGGGCTCGCCACACAAACC	6186
QY	197	LysArgHisSerProAlaLeuGlyArgGlnValValAspMetGlyIleAlaValGlnThr	216
Db	6185	-----TCAACCACAATCTATGATGTGCGCGCAACAGCAGCAACA	6147
QY	217	PheSerAlaLeuAsnValValArg-----Thr	225
Db	6146	GATGCAGCCCTTTCACAGCCCGTCCATCAACAGCAGCGCATGCAATTGCAGCAGCAACA	6087
QY	226	ValLeuAlaProAlaLeuAlaSerArgProSerValGlnGly-----	239
Db	6086	ACAATTGGCCCAACAGCAGCAACAACAGCAGCAGCAGATCCCCAGCATTTTCTCCGA	6027
QY	240	--AlaValAspPheGlyValSerThrAlaGlyGlyLeuValAlaAsnAlaGlyPheGly	258
Db	6026	ATCGCCCAAGATTCTTCAAACTCCACCCATGACAGCAAAAGCTGCATCAGCATCAGGTGGT	5967
QY	259	AspArgMetLeuSerValGlnSerArgAspGlnLeuArgGlyGlyAlaPheValLeuGly	278
Db	5966	-----TCCTGGACAAGCCACTCAGCCGACGCAATCATTTCTCCCA	5928
QY	279	MetLysAspLysGlnProLysAla-----Ala	287
Db	5927	GCAGAAGCCTATCGACCCCAACGAGACCCCGTTCAAGTACCTCAAGTGCTCAGTCGATCTGC	5868
QY	288	LeuSerGlnGluThrAspTrpLeuAspAlaTyrLysAlaIleLysSerAlaSerTyrSer	307
Db	5867	ATTGAGCAGCAATCAGGACTCGTTGATAT-----GAGGACGACAACAATTGAA	5820
QY	308	GlyAlaAlaLeuAsnAlaGlyLysArgMetAlaGlyLeuProLeuAspValAlaThrAsp	327
Db	5819	GCAGCAACAACAGATTCACAACAGCAACAACAATGGCACCAGCCCAAGCAGCAGCAAAAT	5760
QY	328	GlyLeuLysAlaValArgSerLeuValSerAlaThrSerLeuThrLysAsnGlyLeuAla	347
Db	5759	GGCACAACACCCGCAACAGCAGCAGCAGCAACCAACAACAGCAGCACACCCCTTCTCC	5700
QY	348	LeuAlaGlyGlyTyrAlaGlyValSerLysLeuGlnLysMetAlaThr-----	363
Db	5699	GGCACAATCGCCACTGCAACAACGCAACCTACTACCCCACTTGCAACAGCAGCAAAACA	5640
QY	364	-----LysAsnIle-----	366
Db	5639	ACAAAGCGCTCAAGTAAATAAAGAACATATATTTTACCAGAAATTAAATGATTAC	5580
QY	367	-----ThrAspSerAlaThrLysAlaAlaValSerGlnLeu	378
Db	5579	TTCTTTTAGCAAAATACAGCAGCAACAACAACAACAAGCAGCAGCAGCAACAACA	5520
QY	379	SerAsnLeuValGlySerValGlyValPheAla-----GlyTyrPheThrAla	394
Db	5519	ACAGCAGCAGCAACAACAGCAGGTCTCTGACGCGCAGCAACGCCACAGCCTGGACACAGCA	5460
QY	394	-----	394
Db	5459	ACAAGTCATCACTACGCCCATGTTATAAACACATCCACGGCTCAAGGACAACAATCAT	5400
QY	395	-----GlyLeuAlaThrAspProAlaValLysLysAlaGluSerPhe	408
Db	5399	TCAAAGTCATATGTCTTTGGCCTTGCAAGAGCAGCACAATT-----	5358
QY	409	IleGlnAspLysValLysSerThrAlaSerSerThrThrSerTyrValAlaAspGlnThr	428

Accession	Sequence	Annotations
Db 5357	-----GTGCACGCTTCACACACAGCAGCACACAGCCGCAGCAGCACACACAGAT	5307
QY 429	VallysleuAlaLysThrVallysAspMetSerGlyGluAlaIleSerSerThrGlyAla	448
Db 5306	TACTCAAGTGCAGCAACTGCCACCGGCACAGCAGCAGCAACAGCTTCCGCACACACACCA	5247
QY 449	SerLeuArgSerThr 453	
Db 5246	AGTTCACACACACACA 5232	
RESULT 10		
AAAX24980	standard; DNA; 7100 BP.	
AC AAX24980;		
DT 05-JUL-1999	(first entry)	
DE	Bacillus subtilis metalloprotease YOMI DNA.	
KW	YOMI; metalloprotease; protease; textile; animal feed; detergent;	
KW	Gram-positive bacterium; ds.	
OS	Bacillus subtilis.	
XX		
FH	Key	Location/Qualifiers
FT	CDS	112..6969
FT		/*tag= a
XX		
PN	WO914342-A1.	
XX		
PD	25-MAR-1999.	
XX		
PF	08-SEP-1998; 98WO-US18828.	
XX		
PR	15-SEP-1997; 97GB-0019636.	
XX		
PA	(GENV) GENENCOR INT BV.	
PA	(GENV) GENENCOR INT INC.	
XX		
PI	Estell DA;	
XX		
DR	WPI; 1999-229541/19.	
XX	P-PSDB; AAW98149.	
PT	Compositions containing metalloprotease from Gram-positive	
PT	microorganism	
XX		
PS	Claim 25; Fig 1A-O; 59pp; English.	
XX		
CC	This DNA sequence encodes metalloprotease YOMI (see AAW98149) of	
CC	Bacillus subtilis. The invention relates to the discovery of this	
CC	previously unknown metalloprotease, uses of the metalloprotease in	
CC	industrial and agricultural applications, and advantageous strain	
CC	improvements based on genetically engineering a Gram-positive	
CC	microorganism to delete, underexpress or overexpress the enzyme.	
CC	Due to overall relatedness of the enzyme with Pseudomonas lasa	
CC	protein, YOMI appears to be a member of the M23 metalloprotease	
CC	family. The metalloprotease can be used in claimed cleaning	
CC	compositions, animal feed and compositions for the treatment of	
CC	textiles. It may also be used for peptide hydrolysis, waste	
CC	treatment and for cleaving recombinant fusion proteins. Nucleic	
CC	acid sequences are useful as primers and probes for detecting	
CC	similar sequences in other organisms, and for recombinant	
CC	production of the enzyme. Gram-positive cells in which YOMI is	
CC	inductated (by gene mutation or deletion) are used for production	
CC	of heterologous proteins, especially enzymes, hormones, growth	
CC	factors and cytokines.	
XX		
SQ	Sequence 7100 BP; 2649 A; 1294 C; 1353 G; 1804 T; 0 other;	

Pred. No.:	0.139	Length:	7100
Score:	140.00	Matches:	128
Percent Similarity:	33.44%	Conservative:	81
Best Local Similarity:	20.48%	Mismatches:	222
Query Match:	5.82%	Indels:	194
DB:	20	Gaps:	26

US-09-825-414-66 (1-487) x AAX24980 (1-7100)

QY	15	MetGluSerPheArgThrAlaSerAspAlaSerLeuAlaSerSerValArgSerVal	34
Db	862	ATTGAATTAATATACGCCAGCAGCACAAAGTAAATGTTCAAAACCTAAATATACACGGTATGGC	921
QY	35	SerThrTrpSerCysArgAspLeuGlnAlaIleThrAspTyrLeuLysHisValPhe	54
Db	922	AGTTCTATGGCTCTAGTAATAGACAAGCTGTTCAGAATTATTGGAAT-----	969
QY	55	AlaAlaHisArgPheSerVal--IleGlySerProAspGluArgAspAlaIleAla	73
Db	970	GCAGTAATAAGTCTTAATGTAAAGCACTGGAAGCAATAAATATCATGATCACAATTCAAAGC	1029
QY	74	HisAsnGluGlnIleAspAlaLeu-----ValGluThrArgAlaAsnArgLeuTyr	90
Db	1030	TTGAAATATGCAATTTAGAGAATTAGCCCTCCAACGGCTCAACACAGCTGCTAATCAAGCCTCT	1089
QY	91	SerGluGlyGluThrProAlaThrIleAlaGluThrPhe-----	103
Db	1090	TCTTTTGA-----GCAGACTAACCCAAACCCTTCAAAGAAGCATGTCCACCTATTTA	1140
QY	104	-----alalysala	106
Db	1141	ATCTCCGGTCTTTATTCTACGGAGCTATCTTGACCTTAAAGAATGGTATCCACAGCA	1200
QY	107	GluLysPheAspArgLeuAlaThrThrAlaSerSerAlaPheGluAsnThrProPheAla	126
Db	1201	ATAGAAATTGATACTCTCATGACAAATATTCGCCGTGTATGAATGAGCCCGGATTATAAA	1260
QY	127	AlaAlaSerValLeuGlnTyrMetGlnProAlaIleAsnLysGlyAspTrpLeuAlaThr	146
Db	1261	TATAATGAACCTTCTC-----CAAGAATCTAATGACTTAGGTATACACTTCAAAAT	1311
QY	147	ProLeuLysProLeuThrProLeuIleSerGlyAlaLeuSerGlyAlaMetAspGlnVal	166
Db	1312	AAATCACAGATATCTTCAAAATGACAGCGGCAATTTGGAGAAATGGTTTCGATGAAGAAGT	1371
QY	167	GlyThrLysMetMetAspArgAlaArgGlyAspLeuHisTyrLeuSer--ThrSerPro	185
Db	1372	GAGCTCTCCAGCTTAACGAAAACTGCCCAAGTCTTCAAAATGTCTCTGATTAACTCCC	1431
QY	186	AspLysLeuHisAspAlaMetAlaValSerValLysArgHisSerProAlaLeuGlyArg	205
Db	1432	GATGATACAGTTAACTCTTAACGGCAGCAATGCTCAACTTTAATATTCAGCAAAATGAT	1491
QY	206	GlnVal-----ValAspMetGlyIleAlaValGlnThrPhe	217
Db	1492	TCAATATCAATTCGAGATAAATTAAATGAGGTTGATAATACTATGTCTGTATCAACTCTA	1551
QY	218	SerAlaLeuAsnValValArgThrValLeuAlaProAlaLeuAlaSerArgProSerVal	237
Db	1552	GATCTGGCCAATCTATCCGTAAA-----GCTGTTCAACTGCTTCTACATTCGGGGTA	1605
QY	238	GlnGlyAlaValAspPheGlyValSerThrAla-----Gly	249
Db	1606	GAGCTAAATGATCTTATTTGTTATACAACTGCAATTGCTAGTACACACAGTGAATCAGGG	1665
QY	250	GlyLeuValAlaAsn-----AlaGlyPheGlyAspArgMetLeuSer	263
Db	1666	AATATCGTCGGGAACTCCTTAAAGACAATTTTCGGCGCGGATTTGGGAATATCAAACTCA	1725
QY	264	ValGlnSerArgAspGlnLeuArgGlyGlyAlaPheValLeuGlyMetLysAspLysGlu	283
Db	1726	ATTAAGCGCTTAGACACAGATTGGTATCTCAGTTAAAAACAGCTGGTGGTGAAGCTTAAATCA	1785

QY	284	PIOLYSALAALeUSeRglu-----GluThrAspTrileu--AspAlaTyrLys	299
		::: :::	
Db	1786	GCAAGTGAATTAAATTAGTGAAGTTGCTGCTAAGTGCGATACGCTTCTGATGCTCAGAAA	1845
QY	300	AlaIleLysSerAlaSerTyrSerGly-----	308
		:::	
Db	1846	CAAAATACTTCATTTGGAGTAGCTGGTAATTATCATTTATCCCGTTTAATGCAATGATG	1905
QY	309	-----AlaAlaLeuAsnAlaGlyLys-----	315
Db	1906	AACAACCTTCTCTATTGCTCAGAATGCGGCTAAAACTCGCGCTAACTCAACAGAACTGCT	1965
QY	315	-----	315
Db	1966	TGCACTGAGCAGCAAAAGTATGCAGATAGTCTACAAGCTAGGTAATAAGCTTCAAAAT	2025
QY	315	-----	315
Db	2026	AACTTCACTGAATTTGCTATTGACAGCTTGATGCTTTTATTAGCGACGAGTAATTGAA	2085
QY	316	-----ArgMetAlaGlyLeuProLeuAspValAlaThrAspGlyLeuLysAlaVal---	332
		::: ::: ::::: :::::	
Db	2086	TTTACTCAAGCCGACAGGTTCTTGGCTTAACGCTTCTACAGAGATAATCAAAATCAGTTGGG	2145
QY	333	-----ArgSerLeuValSerAlaThrSer-----LeuThrLysAsnGly	345
		:::	
Db	2146	TTTCCTACCTCCCTTTTACGCTGCAGTACGACACTGCAACCTTTTGCCTCAGTAAGATACC	2205
QY	346	LeuAlaLeuAlaGly-----GlyTyrAlaGlyValSerLysLeuGlnLysMet	361
		::: ::: ::: ::: :::	
Db	2206	CGCACATTAGCCACGACGACCTAATTTTGGGCACACGTCGAATGGGGCAA---GAACTTTA	2262
QY	362	AlaThrLysAsnIleThrAspSerAlaThrLysAlaAlaValSerGln-----	377
		::: ::: :::	
Db	2263	GCGACTGCTGGGCTAGAAAGCTGGTATGACTCGTGCAGCAGCTCGCCCTCAAGAGTTCTAAA	2322
QY	378	-----LeuSerAsnLeuValGlySerValGlyValPheAlaGly	390
		:::	
Db	2323	ACTGCTCTTCGAGGGTTGCTTGTTCACTTTAGTTGGCGGTGCATTTGCTGTTGGGA	2382
QY	391	TrpThrThrAlaGlyLeuAlaThrAspProAla---ValLysLysAla-----	405
		:::	
Db	2383	TGGCGGCTAGAAATCATTAATTTCTTTTTCAGAAAGCTAAAAAAGCTAAAGATGATTTT	2442
QY	406	-----	407
		:::	
Db	2443	GAGCAGAGCCAGCAACCAATGTGAGCAATTACGACCATAAAGACTCCACTGATATAA	2502
QY	408	PheIleGln-----AspLysValLysSerThrAlaSerSerThrThrSer	422
		::: :::	
Db	2503	CTAATACAGCAATATAAAGAGCTTCAAAAAGTTAAAGAGTCAAGATCTTTAACTTCAGAT	2562
QY	423	TyrValAlaAspGln-----ThrValLysLeuAlaLysThr-----Val	435
		::: ::::	
Db	2563	GAAAGCAAGAATATACCTTCAAGTCACTCAGCAATTAGACACAAACTTCCCTGCATTAGTT	2622
QY	436	Lys-----AspMetSerGlyGluAlaIleSerSerThrGlyAlaSerLeuArgSerThr	453
Db	2623	AAAGGCTATGATTTCTCAAGGAATGCAATTCTTAAGACAAATAAAGAGCTTGA AAAAGCG	2682
QY	454	ValAsnAsnLeuArg	458
		::: :::	
Db	2683	ATTGAGATACTATAA	2697
RESULT 11			
ABL29839			
ID	ABL29839 standard; DNA; 3690 BP.		
XX	ABL29839;		
AC			
XX			
XX	26-MAR-2002 (first entry)		
DT			
XX			
DE	Drosophila melanogaster genomic polynucleotide SEQ ID NO 40990.		

Drosophila melanogaster genomic polynucleotide SEQ ID NO 40990.

OY 185 ProAspLysLeuHisAspAlaMetAlaValSerValLysArgHis----- 199
 ||||: ||| ||| :::
Db 1492 AACGATGACGAGATCCAAACTCCCGCAGCCAAAGGTCTGGAAATACACGCCAACCATGCTG 1551
OY 200 SerProAlaLeuGlyArgGlnValAlaSpMetGlyLe-----AlaValGlnThr 216
 :::|||| ||| ::: :::
Db 1552 GCACCACTTCTTCCAGGTGAAGAAATGAGTTACAGCGTTGAGCGCTCCGACGTGACAAAC 1611
OY 217 PheSerAlaLeuAsnValValArgThrValLeuAlaProAlaLeuAlaSerArgProSer 236
 ::: ||| ||||| ||| :::
Db 1612 CGCCACAGAAATGGAGAGGTCCGCAACGTTACCGCA-----ACC 1650
OY 237 ValGlnGlyAlaValAspPheGlyValSerThrAlaGlyGlyLeuValAla-----Asn 254
 ||| ||| ||| ||| ||| :::
Db 1651 GTCAACGGCACTTA-----GTGCTTACCGCCACCGCTGTGTGGCAGCTCCATCT 1701
OY 255 AlaGlyPheGlyAspArgMetLeuSerValGlnSerArgAspGlnLeuArgGlyAla 274
 ::: ||| ||| ||| ||| :::
Db 1702 ACTTCTACGCATTCACAGGCCAGGCCATTCAGTCCACAGGC----- 1743
OY 275 PheValLeuGlyMetLysAspLysGluProLysAlaAlaLeuSerGluGluThrAspTrp 294
 :::||||| ::: ||| ||| ||| |||
Db 1744 -----ATGGGTATGGAGCACGCCGTAACCTCTACAGCAGCTCGCGCTATCTGGACCGC 1797
OY 295 LeuAspAlaTyrLysAlaIleLysSerAlaSerTyrSerGlyAlaAlaLeuAsnAlaGly 314
 ||||| ||| ||| ||| ||| |||
Db 1798 GCCGATGCACACACCCGCAATAAG-----CTGGGCTTCTCCATCGTG 1839
OY 315 LysArgMetAlaGlyLeuProLeuAspValAlaThrAspGlyLeuLysAlaValArg--- 333
 ::: ||| :::|||| ||| ||| |||
Db 1840 GAAATCGTGAAACAAACACCACGCGAGTAACCGTGCGAGGGAAGATTCTTCCACCCA 1899
OY 334 ---SerLeuValSerAlaThrSerLeuThrLysAsnGlyLeuAlaLeuAlaGlyGlyTyr 352
 :::|||| ||| ||| ||| ||| |||
Db 1900 GACGGCGTTTGTACCTCACCCAGTTCAACCCAGGTGGCATGGCAACTCTG--GGCGTT 1956
OY 353 AlaGlyValSerLysLeuGlnLysMetAlaThrLysAsnIleThrAspSerAlaThrLys 372
 ||| :::|||| ||| ||| ||| |||
Db 1957 GCTCAGATCGCTGAATGCGTGAAGCACATGCCCTTGAAACCAGCGTCATCTTGCTGGA 2016
OY 373 AlaAlaValSerGlnLeuSerAsnLeuValGlySerValGlyValPheAlaGlyTrpThr 392
 :::|||| ::: ||| ||| ||| |||
Db 2017 CACTCCGTTGTGAGTACACACGCGCTGTGCATATGTGCTGTGTGCTGCCCTGGAATCC 2076
OY 393 Thr-----AlaGlyLeuAlaThrAspProAlaValLysLysAlaGlu 406
 ||| ||| ||| ||| ||| |||
Db 2077 GTTCTGGAGATCGTTTACCGCTGTGCTTGACCATGACACCGCTTGCTGATCGCATGAA 2136
OY 407 SerPheIleGln-----AspLysValLysSerThrAla 417
 ::: ||| ||| ||| ||| |||
Db 2137 AACGTCGTCCAACTACGCGCTCGACGCTCTTCGCCCAACAAGATGGGCTCGACCGCA 2196
OY 418 SerSerThrThrSerTyrValAlaAspGlnThrValLysLeuAlaLysThrValLysAsp 437
 ::: ||| ||| ||| ||| |||
Db 2197 GACAACGTTTTCGATTACGTGCG-----TCTGTTCGGA 2232
OY 438 MetSerGlyGlu-----AlaIle 443
 ||||| ||| ||| ||| ||| |||
Db 2233 GCTTCGGTGAATTCCTGGAGATCGTTAACTACAACCTTGCGCTGGCTGCAGTACGACGTT 2292
OY 444 SerSerThrGlyAlaSerLeuArgSerThrValAsnAsnLeuArgHisArgSerAlaPro 463
 ::: ||| ||| ||| ||| |||
Db 2293 GCTGGAACCCAGGCTGCTCT-----GCCGCCCTT 2322
OY 464 GluAlaAspIleGluGlu-----GlyGlyIleSerAlaPhe-----SerArgSer 478
 ||||| ||| ||| ||| ||| |||
Db 2323 CGTGCCGATTTGAGAACCGTGACACCAAGTCAAGCGTGCCTTCATTTTGATCCCTGCGATT 2382
OY 479 GluThrProPheGlnLeuArgArgLeu 487
 ::: ||| ||| ||| ||| |||
Db 2383 GACGTGCCATTCCACTCCTCCAAGCTG 2409

RESULT 13

AAAF67767
ID AAF67767 standard; DNA; 6136 BP.
XX
AC AAF67767;
XX
DT 11-Apr-2001 (first entry)
XX
DE Corynebacterium glutamicum MCT protein encoding DNA SEQ ID NO:49.
XX
KW Corynebacterium glutamicum; brevibacterium lactofermentum; MCT;
KW membrane construction and membrane transport protein; petroleum spill;
KW hydrocarbon degradation; gram positive aerobic bacterium; marker;
KW identification; microorganism; fine chemical production; transformation;
KW genome mapping; genetic engineering; ds.
XX
OS Corynebacterium glutamicum.
XX
PN WO200100805-A2.
XX
PD 04-JAN-2001.
XX
PE 23-JUN-2000; 2000WO-IB00926.
XX
PR 25-JUN-1999; 99US-0141031.
PR 08-JUL-1999; 99DE-1031454.
PR 08-JUL-1999; 99DE-1031478.
PR 08-JUL-1999; 99DE-1031563.
PR 09-JUL-1999; 99DE-1032122.
PR 09-JUL-1999; 99DE-1032124.
PR 09-JUL-1999; 99DE-1032125.
PR 09-JUL-1999; 99DE-1032128.
PR 09-JUL-1999; 99DE-1032180.
PR 09-JUL-1999; 99DE-1032182.
PR 09-JUL-1999; 99DE-1032191.
PR 09-JUL-1999; 99DE-1032209.
PR 09-JUL-1999; 99DE-1032212.
PR 09-JUL-1999; 99DE-1032227.
PR 09-JUL-1999; 99DE-1032228.
PR 09-JUL-1999; 99DE-1032229.
PR 09-JUL-1999; 99DE-1032230.
PR 14-JUL-1999; 99DE-1032927.
PR 14-JUL-1999; 99DE-1033005.
PR 14-JUL-1999; 99DE-1033006.
PR 27-AUG-1999; 99DE-1040764.
PR 27-AUG-1999; 99DE-1040765.
PR 27-AUG-1999; 99DE-1040766.
PR 27-AUG-1999; 99DE-1040830.
PR 27-AUG-1999; 99DE-1040831.
PR 27-AUG-1999; 99DE-1040832.
PR 27-AUG-1999; 99DE-1040833.
PR 31-AUG-1999; 99DE-1041378.
PR 31-AUG-1999; 99DE-1041379.
PR 31-AUG-1999; 99DE-1041395.
PR 03-SEP-1999; 99DE-1042077.
PR 03-SEP-1999; 99DE-1042078.
PR 03-SEP-1999; 99DE-1042079.
PR 03-SEP-1999; 99DE-1042088.
XX
PA (BADI) BASF AG.
XX
PI Pompejus M, Kroegeer B, Schroeder H, Zelder O, Haberhauer G;
XX
DR WPI; 2001-071486/08.
DR P-PsDB; AAB76534.
XX
PT Corynebacterium glutamicum nucleic acids encoding membrane construction
PT and membrane transport proteins or their portions, useful for typing or
PT identifying C. glutamicum or related bacteria, and as markers for
PT transformation -
XX
PS Claim 3; Page 216-224; 1119pp; English.
XX

KW pharmaceutical; gene; ds.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US09231.
XX
PR 23-MAR-2000; 2000US-191637P.
PR 11-JUL-2000; 2000US-0614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signalling and cell-cell
PT interactions -
XX
PS Claim 1; SEQ ID NO 40987; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA
CC sequences (AB101840-AB16175) and the encoded proteins
CC (AB57737-AB572072).
CC The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 6169 BP; 1607 A; 1369 C; 1664 G; 1529 T; 0 other;

Alignment Scores:
Pred. No.: 0.172 Length: 6169
Score: 138.00 Matches: 107
Percent Similarity: 35.09% Conservative: 73
Best Local Similarity: 20.86% Mismatches: 205
Query Match: 5.73% Indels: 128
DB: 23 Gaps: 21

US-09-825-414-66 (1-487) x ABL29838 (1-6169)

QY 11 ProGlyValAlaMetGluSerPheArgThrAlaSerAspAlaSerLeuAlaSerSerSer 30
Db |||||::: ::|||::: ||| ::|||::: |||
Db 4325 CCCGGAATCACTAAAGTGGCGACATACACCTCGCCGGATATAGTTACAGTTCCGCAAGT 4266

QY 31 -----ValArgSerValSerThrThrSerCysArgAspLeuGlnAlaIleThrAspTyr 48
Db ::|||::: |||::: ||| ::|||::: |||
Db 4265 CCGGGAATATCCAAAGTGGCGACTTACTCTCCCATCGCTTCCAGCGTTCCCTACTAT 4206

QY 49 LeuLysHisHisValPheAlaAlaHisArgPheSerValIleGlySerProAspGluArg 68
Db ::|||::: |||::: ||| ::|||::: |||
Db 4205 GCGCCACCACTGACTTCCAAAGTGGAGACATACAGC-----TCACCT----- 4164

QY 69 AspAlaAlaLeuAlaHisAsnGluGlnIleAsp-----AlaLeuValGluThrArgAla 86
Db ||| ::|||::: ::|||::: ||| ::|||::: |||
Db 4163 -----GCGTACACGTTACAGCAAGACACACACCGGCTACTCCAAAGGTGGAGACCTACAGC 4110

QY 87 AsnArgLeuTyrSerGluGlyGluThrProAlaThrIle----- 99
Db ::|||::: |||::: ||| ::|||::: |||
Db 4109 TCTCTGTTACAGCTATGACAGATCTCGCCGGCATCTCCAGAGATAGCCACCTACTCG 4050

QY 100 -----AlaGluThrPheAlaLysAlaGluLysPheAsp----- 110
Db ||| ||| ||| ||| ||| :::
Db 4049 CCATCGGTTTCGATTTCAGCGCCCAACAGATTGCCAAGGTGCCACCTACTCAGCGCTTCA 3990

QY 111 ---ArgLeuAlaThrThrAlaSer---SerAlaPheGluAsnThrProPheAlaAlaAla 128
Db ::|||::: |||::: ||| ::|||::: |||
Db 3989 GTGAACCTGGCCACACTTCCTGCTACTGCTCTCTCCACCGAAGACAGATACTCCGCCAGC 3930

QY 129 -----SerValLeuGlnTyrMetGlnProAlaIleAsnLysGlyAspTyrLeuAla 145
Db |||::: |||::: ||| ::|||::: |||
Db 3929 TATGCTCCCTCAATTACCAAGTACAGCAGGATGATGATATCCACCAAGTACATATCC 3870

QY 146 ThrProLeuLysProLeuThrProLeuIleSer-----GlyAla 158
Db |||::: |||::: ||| ::|||::: |||
Db 3869 AAGCCTATAGTGGCAGCCTATCCGGCATCACCAAGTGGCGCCAGCTACGAGGCACT 3810

QY 159 LeuSerGlyAlaMetAspGln-----ValGlyThrLysMetMetAsp 172
Db |||||::: |||::: ||| ::|||::: |||
Db 3809 GCATCCGAGCGCTGTCCACCAAGTATGTCTCCCAACCAAGCATTGCCCAAGTGTCCACT 3750

QY 173 ArgAlaArgGlyAspLeuHisTyrLeuSerThrSerPro-----AspLysLeuHisAsp 190
Db ||| ::|||::: |||::: ||| ::|||::: |||
Db 3749 TACGCTGCTCCACCGGTGCAACTTACTCATCTGGGCTGCCATCTCGAAGCTCTCCACC 3690

QY 191 AlaMetAlaValSerValLysArgHisSerProAlaLeuGlyArgGlnValAlaAspMet 210
Db ::|||::: |||::: ||| ::|||::: |||
Db 3689 AGCTATGGAGCTTCA-----GGATCGGAGCCGTCTCCACCAATATGTCTCGAAG 3639

QY 211 GlyIleAlaValGlnThrPheSerAlaLeuAsnValValArgThrValLeuAlaProAla 230
Db |||::: |||::: ||| ::|||::: |||
Db 3638 CCAGCTGTGCTATTGCCGCTCCACAGCTGCGCAAGGTGCAACGTATGACGCTCCGCC 3579

QY 231 LeuAlaSer-----ArgProSerValGlnGlyAlaValAspPheGlyValSerThr 247
Db ::|||::: |||::: ||| ::|||::: |||
Db 3578 ATTTCAAGTACTTCCACCTGCTCCGATTCGCAAGGTAGCATCTACGACGCTCCACT 3519

QY 248 AlaGlyGlyLeuValAlaAsnAlaGlyPheGlyAspArgMetLeuSerValGlnSerArg 267
Db ::|||::: |||::: ||| ::|||::: |||
Db 3518 GTCTCCACCTACTCTCGGATATGTTATGTTG----- 3483

QY 268 AspGlnLeuArgGlyGlyAlaPheValLeuGlyMetLysAspLysGluProLysAlaAla 287
Db ||| ::|||::: ||| ::|||::: |||
Db 3482 -----AGTGGGTACAGGGCCGTATCGCACCAAGTACGTGCCAAGCCTGCGGTGCC 3432

QY 288 LeuSerGluGluThrAspTyrPheLeuAspAlaTyrLysAlaIle---LysSerAlaSerTyr 306
Db ::|||::: |||::: ||| ::|||::: |||
Db 3431 ATTTC-----GCCGCTCCACCATCGCCCAAGGTGGCTACTAT 3393

QY 307 SerGlyAlaAlaLeuAsnAlaGlyLysArgMetAlaGlyLeuProLeu----- 322
Db ::|||::: |||::: ||| ::|||::: |||
Db 3392 GCAGCTCCCGCATTTCACT-----TATCGCGACGCTCCAGTGTGACCAAAAGTC 3342

QY 322 ----- 322
Db 3341 GCCACTGGATATGAGAGAGTGGCTCGGATACAGTTCCGGCGCTGTTCCCATCATATAT 3282

QY 323 -----AspValAlaThrAspGlyLeuLysAlaValArgSerLeu 335
Db |||||::: |||::: ||| ::|||::: |||
Db 3281 GTCTCTAAGCCAGCGGTGCTAAAGTGGCAACCTATGCTCTGCCATTTCAACATATAC 3222

QY 336 ValSerAlaThrSerLeuThrLysAsnGlyLeuAlaLeuAlaGlyLysTyrAlaGlyVal 355
Db ::|||::: |||::: ||| ::|||::: |||
Db 3221 TCAGCAGCTCCGGCAGTGCACCAAGATCGGCACAAGCTATGAGGATCGGGTCATGGAAGCC 3162

QY 356 SerLysLeuGlnLysMetAlaThrLysAsnIleThrAspSerAlaThrLysAlaAlaVal 375
Db ||| ::|||::: |||::: ||| ::|||::: |||
Db 3161 GTCTCCCATCAGTATGTCTCCAAAGCCAGCAGTACCATCTCAGCAGCCCGCGCA----- 3108

QY 376 SerGlnLeuSerAsnLeuValGlySerValGlyValPheAlaGlyTyrThrThrAlaGly 395
Db ::|||::: |||::: ||| ::|||::: |||
Db 3107 -----ATCGCTAAGGTAGCGACATACGCTTCTCCGGCAATATCAACG 3066

QY 396 LeuAlaThrAspProAlaValLysLysAlaGluSerPheIleGlnAspLysValLysSer 415
Db ||||| ||| ||| ||| ::|||::: |||
Db 3065 TATGCCACTGCTCTGTGGTCTCAAAAGTGGCTACTACGACAGCTCCCTCCATGAGCCACT 3006

QY 416 ThrAlaSerSerThrThrSerTyrValAlaAspGlnThrValLysLeuAlaLysThr--- 434

Db 3005 TACTCCTCCGCTCCGGCT-----CTAGCCAAGGTTTCC 2973
QY 435 -----VallysAspmetSerGlyAlaIleSer 444
Db 2972 TATAGTCAGGACGGATGTATCCACCATGATTTCC 2934
RESULT 15
AAH67704
ID AAH67704 standard; DNA; 8979 BP.
AC AAH67704;
XX 26-SEP-2001 (first entry)
DE C glutamicum coding sequence fragment SEQ ID NO: 2739.
XX
KM Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
KW organic acid synthesis; ds.
XX
OS Corynebacterium glutamicum.
PN EP1108790-A2.
XX 20-JUN-2001.
PF 18-DEC-2000; 2000EP-0127688.
XX
PR 16-DEC-1999; 99JP-0377484.
PR 07-APR-2000; 2000JP-0159162.
PR 03-AUG-2000; 2000JP-0280988.
XX
PA (KYOW) KYOWA HAKKO KOGYO KK.
XX
PI Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
XX
DR WPI; 2001-376931/40.
DR P-PSDB; AAG92485.
XX
PT Novel polynucleotides derived from Coryneform bacteria, for identifying
PT mutation point of a gene, measuring expression of a gene, analysing
PT expression profile or pattern of a gene and identifying homologous gene
PT
XX
PS Claim 8; SEQ ID NO: 2739; 246bp + Sequence Listing; English.
XX
CC The present invention provides a number of nucleotide and protein
CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These
CC are useful for identifying the mutation point of a gene derived from a
CC mutant of coryneform bacterium, measuring expression amount and
CC analysing the expression profile or expression pattern of a gene derived
CC from Coryneform bacterium, and identifying a homologue of a gene derived
CC from coryneform bacterium. Coryneform bacteria are useful for producing
CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
CC particularly L-lysine. The present sequence is a nucleic acid described
CC in the exemplification of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC European Patent Office.
XX
SQ Sequence 8979 BP; 1749 A; 2763 C; 2544 G; 1923 T; 0 other;

Alignment Scores:
Pred. No.: 0.277 Length: 8979
Score: 138.00 Matches: 119
Percent Similarity: 37.16% Conservative: 85
Best Local Similarity: 21.68% Mismatches: 219
Query Match: 5.73% Indels: 126
DB: 22 Gaps: 23

US-09-825-414-66 (1-487) x AAH67704 (1-8979)

QY 7 AlaGlnGlnProProGlyValAlaMetGluSerPheArgThrAlaSerAspAlaSerLeu 26
Db 3073 GCACACACCGCACCGTGCAGTGTGGCCGCGCATGGCCACGACGATTTTCCTGCCGTA 3132
QY 27 AlaSerSerSerValArgSerValSerThrThrSerCysArgAspLeuGlnAlaIleThr 46
Db 3133 AAGTCCGAGTCAATCCAGGACCGATTCGCCATCCGTTGTGGAAGGCATGCTCCCTG 3192
QY 47 AspTyrLeuLysHisHisValPheAlaAlaHisArgPheSerValIleGlySerProAsp 66
Db 3193 GTTCACCTGGAGCACACATTT-----GTGCTCAAGTCCGATGTC 3231
QY 67 GluArgAspAlaAlaLeuAlaHisAsnGlnIleAspAlaLeuValGluThrArgAla 86
Db 3232 CCAACCGACGCGCGCTGAAGGTTTCCGCGACTGCCGATGAGGTAGTCGATACCGACCTG 3291
QY 87 AsnArgLeuTyr-----SerGluGlyGluThrProAlaThr 98
Db 3292 GGTCCGCTCGTGAATGCTGCGCGCAGAAATCGCCGACGACAGAAAGCAACCTGATGCTACG 3351
QY 99 IleAlaGluThrPheAla---LysAlaGluLysPheAspArgLeuAlaThrThrAlaSer 117
Db 3352 TTGGCTGAGCGTTTCGCGATCCGCGGACGCAAGGCAACGCTGTCCGACGCAACACACC 3411
QY 118 SerAlaPheGlu-----AsnThrProPheAlaAlaAlaSerValLeuGlnTyrMet 134
Db 3412 TCCGCACTGCCAACCAACCGTGGACACCCACGCTCAGCTCGCGCAGTGGCAACCGTTGTT 3471
QY 135 GlnPro-----AlaIleAsnLysGlyAspTyrPleuAlaThrProLeu 148
Db 3472 GCACCTGAATCCATGCGCCCATTTGCTGTGATCTCCGGTGAC----- 3513
QY 149 LysProLeuThrProLeu-----IleSerGlyAlaLeuSerGlyAlaMetAsp 164
Db 3514 -----CGCAACCCATTCACGTCCTGTGATGTGCGGCTTCCCTGGCTGTGCCAGGT 3567
QY 165 GlnValGlyThrLysMetMetAspArgAlaArgGlyAspLeuHisTyrLeuSerThrSer 184
Db 3568 GTGATCGTGCACGGCATGTGGACCTCTGCCATCGGTGAACGTATCGCCGTCACGATTC 3627
QY 185 ProAspLysLeuHisAspAlaMetAlaValSerValLysArgHis----- 199
Db 3628 AACGATGAGCAGATCCAACACTCCCGCAGCCCAAGCTGTGAATACACCCGCAACCATGCTG 3687
QY 200 SerProAlaLeuGlyArgGlnValValAspMetClyIle-----AlaValGlnThr 216
Db 3688 GCACCAAGTTCCTCCACGCTGAAGAATTCAGTTTCAGCCTGAGCGCTCCGCTGACACAAC 3747
QY 217 PheSerAlaLeuAsnValValArgThrValLeuAlaProAlaLeuAlaSerArgProSer 236
Db 3748 CGCCCAAGAAATGGAGAGGTCCGACCCGCTTACCGCA-----ACC 3786
QY 237 ValGlnGlyAlaValAspPheGlyValSerThrAlaGlyGlyLeuValAla-----Asn 254
Db 3787 GTCAACGGCAACTTA-----GTGCTTACCGCCACCGCTGTGTGGCAGCTCCATCT 3837
QY 255 AlaGlyPheGlyAspArgMetLeuSerValGlnSerArgAspGlnLeuArgGlyAla 274
Db 3838 ACTTCTACGCATTCCTCCAGCGCAGGCAATTCAGTCCAGGCG----- 3879
QY 275 PheValLeuGlyMetLysAspLysGluProLysAlaAlaLeuSerGluGluThrAspTyr 294
Db 3880 -----ATGGGTATGGAAGACACGCGCTAATCTCAGGCAAGCTCGCGCTAATCTGGACCGC 3933
QY 295 LeuAspAlaTyrLysAlaIleLysSerAlaSerTyrSerGlyAlaAlaLeuAsnAlaGly 314
Db 3934 GCCGATGCACACACCGCAATTA-----CTGGGCTTCTCCATCTGTG 3975
QY 315 LysArgMetAlaGlyLeuProLeuAspValAlaThrAspGlyLeuLysAlaValArg--- 333
Db 3976 GAAATCGTGAACAAACACCGCAGGAACTAAGTGGCAGGGGAGAAAGTCTTCCACCCA 4035
QY 334 ---SerLeuValSerAlaThrSerLeuThrLysAsnGlyLeuAlaLeuAlaGlyGlyTyr 352

```
Db 4036 GACGGCGTTTGTACCTCACCACAGTTACCCAGGTGGCATGGCACTCTG--GGCGTT 4092
      ::::: ||| ||::: ||::: |||
QY 353 AlaGlyValSerLysLeuGlnLysMetAlaThrLysAsnIleThrAspSerAlaThrLys 372
      ||| ::::: |||
Db 4093 GCTCAGATCGCTGAATGCGTGAAGACATGCGCTTGACCACGCCGTCATCTTGCCTGGA 4152
      ::::: |||
QY 373 AlaAlaValSerGlnLeuSerAsnLeuValGlySerValGlyValPheAlaGlyTrpThr 392
      ::::: |||
Db 4153 CACTCCGTTGGTAGTACACCGCGCTTGTCATATGCTGGTGTGCTGTCCTGGAATCC 4212
      ::::: |||
QY 393 Thr-----AlaGlyLeuAlaThrAspProAlaValLysLysAlaGlu 406
      ||||| ||| ::::: |||
Db 4213 GTTCTGGAGATCGTTTACCGTGGCTGACCATGCACCGCTTGGTGGATCGCGATGAA 4272
      ::::: |||
QY 407 SerPheIleGln-----AspLysValLysSerThrAla 417
      ::::: |||||
Db 4273 AACGGTCTGTCCAACACTACGCGCTCGAGCTCTTCGCCCAACAAGATGGGTCTGACCGCA 4332
      ::::: |||
QY 418 SerSerThrThrSerTyrValAlaAspGlnThrValLysLeuAlaLysThrValLysAsp 437
      ::::: |||||
Db 4333 GACAACGTTTTCGATTACGTGCG-----TCTGTTCCGAA 4368
      ::::: |||
QY 438 MetSerGlyGlu-----AlaIle 443
      ||||| |||
Db 4369 GCTTCGCGTGAATTCTTGAGATCGTTAACTACAACTTGGCTGGCGCTGCAGTACGAGTT 4428
      ::::: |||
QY 444 SerSerThrGlyAlaSerLeuArgSerThrValAsnAsnLeuArgHisArgSerAlaPro 463
      ::::: |||
Db 4429 GCTGGAACCCAGGCTGCTCT-----GCCGCCCTT 4458
      ::::: |||
QY 464 GluAlaAspIleGluGlu-----GlyGlyIleSerAlaPhe-----SerArgSer 478
      ||||| ::::: |||
Db 4459 CGTGCCGATGTTGAGAACCCTGCACAGGTACGCGTGCCTTCATTGATCCCTGCGATT 4518
      ::::: |||
QY 479 GluThrProPheGlnLeuArgArgLeu 487
      ::::: |||||
Db 4519 GACGTGCCATTCCACTCTCCAAGCTG 4545
      ::::: |||
```

Search completed: January 31, 2003, 05:30:03
Job time : 293.246 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: January 31, 2003, 03:21.14 ; Search time 1575.62 seconds
(without alignments)
5005.784 Million cell updates/sec

Title: US-09-825-414-66
Perfect score: 2407
Sequence: 1 MHINQSAQQPPGVAMESFRT.....EEGISAFAFSRSETPFQLRRL 487

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 16154066 segs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODEL=frame+ p2n.model -DEV=xlp
-O=/cgn2_1/USPTO_spool/US09825414/runat_28012003_161742_26093/app_query.fasta_1.1294
-DB=EST -QFMT=fastap -SUFFIX=p2n.rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=1 -MATRIX=biosum62 -TRANS=human40.cdi -LIST=45
-DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09825414@cgn2_1_1_2221_@runat_28012003_161742_26093 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : EST: *
1: em_estba: *
2: em_esthum: *
3: em_estlin: *
4: em_estmu: *
5: em_estov: *
6: em_estpl: *
7: em_estro: *
8: em_htc: *
9: gb_est1: *
10: gb_est2: *
11: gb_htc: *
12: gb_est3: *
13: gb_est4: *
14: gb_est5: *
15: em_estfun: *
16: em_estom: *
17: gb_gss: *
18: em_gss_hum: *
19: em_gss_inv: *
20: em_gss_pln: *
21: em_gss_vrt: *
22: em_gss_fun: *
23: em_gss_mam: *
24: em_gss_mus: *
25: em_gss_other: *
26: em_gss_pro: *
27: em_gss_rod: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	136	5.7	2191	17	AQ012154
C 2	128	5.3	883	12	BG369810
C 3	122	5.1	4001	11	AK014542
C 4	115.5	4.8	700	9	AL508808
C 5	113.5	4.7	1094	17	CNS076CM
C 6	111.5	4.6	833	14	BQ800257
C 7	107	4.4	680	9	AL508002
C 8	107	4.4	842	14	BQ800337
C 9	107	4.4	1150	17	CNS079Z5
C 10	107	4.4	1763	11	AY107053
C 11	105.5	4.4	841	14	BQ800379
C 12	105.5	4.4	850	12	BF265543
C 13	105.5	4.4	888	17	CNS076CP
C 14	105	4.4	579	14	BQ795937
C 15	105	4.4	806	13	BM077210
C 16	105	4.4	853	14	BQ800263
C 17	105	4.4	886	14	BQ794648
C 18	104.5	4.3	842	14	BQ792279
C 19	104.5	4.3	965	14	BQ674550
C 20	104	4.3	587	10	BE583087
C 21	104	4.3	858	12	BF265524
C 22	103.5	4.3	613	9	A1517569
C 23	103.5	4.3	842	14	BQ797946
C 24	103.5	4.3	2779	11	AK004928
C 25	103	4.3	718	14	BQ799997
C 26	103	4.3	743	14	BQ793911
C 27	103	4.3	774	13	B1556804
C 28	103	4.3	778	14	BQ7999265
C 29	103	4.3	800	14	BQ794248
C 30	103	4.3	821	14	BQ796730
C 31	103	4.3	871	14	BQ794592
C 32	103	4.3	933	12	BF065640
C 33	103	4.3	1034	17	CNS0606U
C 34	103	4.3	1068	12	BF540337
C 35	102.5	4.3	2539	11	AK004265
C 36	102	4.2	1009	14	BM904265
C 37	101.5	4.2	681	13	BM582171
C 38	101.5	4.2	989	14	BQ065543
C 39	101.5	4.2	1031	17	CNS06D42
C 40	101.5	4.2	1138	17	CNS060E0
C 41	101	4.2	570	10	AW553832
C 42	101	4.2	691	10	AW553048
C 43	101	4.2	763	17	AQ012153
C 44	100.5	4.2	642	13	B1952036
C 45	100.5	4.2	649	13	B1958467

ALIGNMENTS

RESULT 1
LOCUS AQ012154/c 2191 bp DNA linear GSS 04-JUN-1998
DEFINITION 8758P1D037011098 Cosmid library of chromosome II Rhodobacter
sphaeroides genomic clone 8758P1D037011098, DNA sequence.
ACCESSION AQ012154
VERSION AQ012154
KEYWORDS GI:3177109
SOURCE
ORGANISM Rhodobacter sphaeroides.
Rhodobacter sphaeroides.
Bacteria; Proteobacteria; Alphaproteobacteria; Rhodobacteriales;
Rhodobacteraceae; Rhodobacter.
REFERENCE 1 (bases 1 to 2191)
AUTHORS Choudhary,M., Mackenzie,C., Mouncey,N., Weinstein,G.M. and Kaplan ,S.

sequencing pipeline with 384 multicapillary sequencer
Genome Res. 10 (11), 1757-1771 (2000)

JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS

4

Kawai,J., Shinagawa,A., Shibata,K., Yoshino,M., Itoh,M., Ishii,Y.,
Arakawa,T., Hara,A., Fukunishi,Y., Konno,H., Adachi,J., Fukuda,S.,
Aizawa,K., Izawa,M., Nishi,K., Kiyosawa,H., Kondo,S., Yamanaoka,I.,
Saito,T., Okazaki,Y., Gojobori,T., Bono,H., Kasukawa,T., Saito,R.,
Kadota,K., Matsuda,H., Ashburner,M., Batalov,S., Casavant,T.,
Fleischmann,W., Gaasterland,T., Gissi,C., King,B., Kochiwa,H.,
Kuehl,P., Lewis,S., Matsuo,Y., Nikaido,I., Pesole,G.,
Quackenbush,J., Schriml,L.M., Staabli,F., Suzuki,R., Tomita,M.,
Wagner,L., Washio,T., Sakai,K., Okido,T., Furuno,M., Aono,H.,
Baldarelli,R., Barsh,G., Blake,J., Boffelli,D., Bojunga,N.,
Carninci,P., de Bonaldo,M.F., Brownstein,M.J., Bult,C.,
Fletcher,C., Fujita,M., Gariboldi,M., Gustincich,S., Hill,D.,
Hofmann,M., Hume,D.A., Kamiya,M., Lee,N.H., Lyons,P.,
Marchionni,L., Mashima,J., Mazarrelli,J., Mombaerts,P., Nordone,P.,
Ring,B., Ringwald,M., Rodriguez,I., Sakamoto,N., Sasaki,H.,
Sato,K., Schonbach,C., Seya,T., Shibata,Y., Storch,K.F., Suzuki,H.,
Toyo-Oka,K., Wang,K.H., Weitz,C., Wittaker,C., Wilming,L.,
Wynshaw-Boris,A., Yoshida,K., Hasegawa,Y., Kawai,I.H., Kohtsuki,S.
and Hayashizaki,Y.

Functional annotation of a full-length mouse cDNA collection
Nature 409 (6821), 685-690 (2001)

TITLE
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS

5

(bases 1 to 4001)
Adachi,J., Aizawa,K., Akahira,S., Akimura,T., Aono,H., Arai,A.,
Arakawa,T., Baldarelli,R., Bono,H., Brownstein,M., Bult,C.,
Carninci,P., Fukuda,S., Fukunishi,Y., Furuno,M., Hanagaki,T.,
Hara,A., Hayatsu,N., Hill,D., Hiramoto,K., Hiraoka,T., Hori,F.,
Hume,D., Imotani,K., Ishii,Y., Itoh,M., Izawa,M., Kasukawa,T.,
Kato,H., Kawai,J., Kojima,Y., Konno,H., Kouda,M., Koya,S.,
Kurihara,C., Matsuyama,T., Miyazaki,A., Nishi,K., Nomura,K.,
Numazaki,R., Ohno,M., Okazaki,Y., Okido,T., Owa,C., Quackenbush,J.,
Saito,H., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki,D.,
Schriml,L., Shibata,K., Shibata,Y., Shinagawa,A., Shiraki,T.,
Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F.,
Tanaka,T., Tejima,Y., Toya,T., Yamamura,T., Yamanaoka,I.,
Yasunishi,A., Yoshida,K., Yoshino,M., Muramatsu,M. and
Hayashizaki,Y.

Direct Submission
Submitted (10-JUL-2000) Yoshinide Hayashizaki, The Institute of
Physical and Chemical Research (RIKEN), Laboratory for Genome
Exploration Research Group, RIKEN Gemomic Sciences Center (GSC),
RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
Fax:81-45-503-9216)

COMMENT
Please visit our web site (http://genome.gsc.riken.go.jp/) for
further details.

CDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues. First strand cDNA was primed with a primer
[5', GAGAGAGAGAAGGATCCAGAGAGCTCTTTTCTTTTNN 3'], cDNA was
prepared by using trehalose thermo-activated reverse transcriptase
and subsequently enriched for full-length by cap-trapper. cDNA went
through one round of normalization to Rot = 10.0 and subtraction to
Rot = 100.0. Second strand cDNA was prepared with the primer
adapter of sequence [5',
GAGAGAGAGATTCGAGTTAATTAAATTAAATCCCCCCCCCCC 3']. cDNA was cleaved
with BamHI and XhoI. cDNA of size comprised longer than 7 kb was
selected before cloning. Vector: a modified pluescript KS(+) after
bulk excision from lambda FLC I. Cloning sites, 5' end: SalI; 3'
end: BamHI. Host: DH10B.

FEATURES
source
1. 4001
/organism="Mus musculus"
/strain="C57BL/6J"

/db_xref="FANTOM_DB:4631434016"
/db_xref="MGD:MGI:1897732"
/db_xref="taxon:10090"
/clone="4631434016"
/tissue_type="skin"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="0 day neonate"
1. 4001
/gene="Hiplr"
114. 3320
/gene="Hiplr"
/note="data source:MGD, source key:MGI:1352504,
evidence:ISS
huntingtin interacting protein 1 related
putative"
/codon_start=1
/protein_id="BAB29420.1"
/db_xref="GI:12852463"
/db_xref="MGD:MGI:1352504"
/translation="MNSIKNVPARVLSRTPGHSLAEERQFDKTOAISISKAINSGEA
PYEKHARRIILGTHHEKGFYWSYAIGPLSSSSISMKFCHVLHKVLRDGHNVLA
HDYQRRSNIREFIGDLWGHDYGHVNIYTKILTKFSHLKHPOFPAGLEVPEV
LEKAAGTDVNNIFOLTVEEMFDYMDCELKLSSEVFROLNTAIVSOMSSGCCRLAPLIQ
VIQDCHLYHTVTKLMFKHSCLPADTLQGRDRFHEDHSLKNFERASDMLYFKRL
IQIPRLPEGPNFLRASALAEHIKPVVPIPEAPEEEEPENLEISSAPPAGEVYVA
DLFDQTFGPNGSMKDDRDLQIENLKREVETLRALERIKMEAQRYISQLKGQVNGLE
AELEQROKOKALVDNEQLRHELAQLKALOLEGARNGVLSRAALDSVSGLEOHTOYLAAS
KEKHSSELINTHAEILLRNADTAQKOLTVTQOQSEEVARYKEQLAFOMEQAKRESEKME
EQQDLEKRLRELARAGELARQAEALRTEQSGSELSRLDTLNAEKELSGVVROR
EAEILAAQSLVREKEALSDQQRSSQSEKELRGQLAKRESQEOGLRQKLLDEQLAVL
RSAAEAEAILQDAVSKLDDPLRLCTSSPDYLSRAALDSVSGLEOHTOYLAAS
EDASALVAALTRESHLADTYVNGAATSHLAPTPDPAIDMDTCREGARALELVQQLQ
DQTVLPRAQPSLMRAPLQIGILOQDILKPKSLDVRQELGAMVDKEMATSAATEDAY
RRIDEMMSQARHESSGVKLENERILNSCTDLMKAIRLVTSTSLQKEIVESGRGA
TQOEYVAKNSRWTEGLISAKAVGWGATQLVESADKAVLHMGKYEBELIVCSHEIAST
AOLVAASKVANKNSPHLSRQECSTRVNERAANVAVSTRSGEOIEDRTMDFSGLS
LIRIKQEMETQVRVLELKTLEAERVRIGELRKOHYVLAGMGTPSEEPSPSPAP
RSGATRKPPPLAOKPSIAPRFDNDKKGDVYPAQLVNY"

BASE COUNT 925 a 1120 c 1260 g 695 t 1 others
ORIGIN

Alignment Scores:
Pred. No.: 0.0737 Length: 4001
Score: 122.00 Matches: 120
Percent Similarity: 37.26% Conservative: 73
Best Local Similarity: 23.17% Mismatches: 185
Query Match: 5.07% Indels: 141
DB: 11 Gaps: 26

US-09-825-414-66 (1-487) x AK014542 (1-4001)

QY 1 MethisIleasnGlnSerAlaGlnGlnProProGlyValAlaMetGlnSerPheArgThr 20
DB 1626 ATGAAGATGGAAGACGACGAGCAGTGTGGAAGAAGCTCAAGAGGAGCTGGCCGACAG 1685
QY 21 AlaserAspAlaserLeuAlaserSerSerValArgSerValSerThrThrSerCysArg 40
DB 1686 GCAGAGAGAGCTGGCCGTCGCACGAGGCCCTG--AGCCGCACAGAAACAGAGTGGTCA 1742
QY 41 AspleuGlnAlaIleThrAspTyrLeuLysHisValPheAlaAlaHisArgPheSer 60
DB 1743 GAGCTGAGCTCAGCGCTGACACACTGAAT-----GCCGAGAAGAAGACT 1787
QY 61 ValIleGlySerProAspGluArgAspAlaAlaLeuAlaHis-----Asn 75
DB 1788 CTGAGTGGAGTCGTTCCGACGCGTGAGGCGACAGCTGCTGGCCGCTCAGACCTGGTGGG 1847
QY 76 GluGlnIleAspAlaLeuValGluThrArgAlaAsnArgLeuTyrSerGluGlyGluThr 95
DB 1848 GAGAAGGAGGAGGACCTTAGCCAAAGAGACGACGCGAGAGCTCCACAGAGAGGCGAGCTA 1907
QY 96 ProAlaThrIleAlaGluThrPheAlaLysAlaGluLysPhe----- 109

Db 1908 CGGGGGCAGCTGGCAGAAAAGAGTCTCAGAGCAGGGGCTTCGGCAGAGCTGCTGGAT 1967

QY 110 AspargLeuAlaThrThrAlaSerSerAlaPheGluAsnThrProPheAlaAlaSer 129

Db 1968 GAGCAGTTGGCGGTGTGGCAAGT-----GCAGCCGGCCGAG 2003

QY 130 ValLeuGlnTyrMetGlnProAlaIleAsnLys---GlyAspTrpLeuAlaThrProLeu 148

Db 2004 GCAGAGGCCATCCTACAGATGCAGTGAGCAGAGCTGGACGACCC-CCTGCACCTCCGCTG 2062

QY 149 LysProLeuThrProLeuIleSerGlyAlaLeuSerGly-AlaMetAspGlnValGlyTh 168

Db 2063 CACCAGCTCCCGAGACTACTGTGAGCCGGGCTCAGGCGACCCCTGGACAGCGTGAGC-- 2120

QY 168 rLysMetMetAspArgAlaArgGlyAspLeuHisTyrLeuSerThrSerProAspLysLe 188

Db 2121 -----GGCCTGGAGCAGGGCCACACCAGCTACCTGGCTTCTGTAAGATGCT-- 2168

QY 188 uHisAspAlaMetAlaValSerValLysArgHisSerProAlaLeuGlyArgGlnValVa 208

Db 2169 ----TCTGCCCTGGTGGCAGCGCTGACCCGCTTCTCCATTGTGGCTGCGACACCAATTGT 2224

QY 208 lAspMetGlyIleAlaValGlnThrPheSer----- 218

Db 2225 CAAT---GGTGCCGCCACCTCCACCTGGCCCCCACCAGACCCCGCGCTGATGA 2281

QY 219 -----AlaLeuAsnValVal-----ArgTh 225

Db 2282 CACATGCAGGAGTGTGAGCCCGGCTCTGGAGCTGTGGGACAGCTGCACAGACCAGAC 2341

QY 225 rValLeuAlaProAlaLeuAlaSer-----ArgProSerValGlnGlyAlaValAspPh 243

Db 2342 AGTGCTACCGAGGGCTCAGCCCAAGCTGATGGGGCCCCCTGCACAGGGCATTTCTGCATT 2401

QY 243 eGlyValSerThrAlaGlyGlyLeuValAlaAsnAlaGlyPheGlyAspArgMetLeuSe 263

Db 2402 GGGCCAGGACTG-----AAGCCTAAGAG 2425

QY 263 rValGlnSerArgAspGlnLeuArgGlyGlyAlaPheValLeuGlyMetLysAspLysG1 283

Db 2426 CTTGATGTACGGCAAGAGAGCTAGGG-----GCCATGGTGACACAGA 2470

QY 283 uProLysAlaAlaLeuSerGlnGluThrAspTrpLeuAspAlaTyrLysAlaIleLysSe 303

Db 2471 G---ATGGCGGCACCTCGCAGCCATTGAG-----GACGCTGTGCGGAGATGACAGA 2521

QY 303 rAla-----SerTyrSerGlyAlaAlaLeuAsnAlaGlyLysArgMe 317

Db 2522 CATGATGAGCAGGCCCGCCAGCAGAGAGCTCGGGCGTGAACCTGAGGTGATGAGAGCAT 2581

QY 317 talagLysLeuProLeuAspValAlaThrAspGlyLeuLysAlaValArgSerLeuVal-- 336

Db 2582 C-----CTCACTTCCTGCACAGACCTGATGAAGGCTATCCGGCTCTGTGAT 2629

QY 337 -SerAlaThrSerLeuThrLysAsnGlyLeuAlaLeuAlaGlyGlyTyrAlaGlyValSe 356

Db 2630 GACCTCCACAGCCTGCAGAGGAA-----ATTGTGAGAGCGGCGAGGGGCGAC 2680

QY 356 rLysLeuGlnLysMetAlaThrLysAsnIle-----ThrAsp-----SerAlaTh 371

Db 2681 AACGCAGCAGGAATTTTATGCCAAGAATTACAGGTGACTGAAGCCCTCATCTCAGCCTC 2740

QY 371 rLysAlaAlaValSerGlnLeuSerAsnLeuValGlySerValGlyValPheAlaGlyTr 391

Db 2741 TAAGGCA-----GTGGGCTG 2755

QY 391 pThrThrAlaGlyLeuAlaThrAspProAlaValLysLysAlaGlySerPheIleGlnAs 411

Db 2756 GGGAGCCACACAGCTG-----GTGGAGTCAGCTGCACAAGGTTGTGCTTCA 2800

QY 411 pLysValLys-----SerThrAlaSerSerThrThrSe 422

Db 2801 CATGGGCAAAATAGAGGAACATCATGCTGTCTCCATGAGATTTGGGCCAGCAGCGCCCA 2860

QY 422 rTyrValAlaAspGlnThrValLys----- 430

Db 2861 GCTGCTGGCAGCCTCGAAGGTGAAGCCACACAGACAGTCCCACTTGAGCCGCTGCA 2920

QY 431 ----LeuAlaLysThrValLysAspMetSerGlyGluAlaIleSerThr 446

Db 2921 GGAATGTTCCCGCACTGTCAAGCAGAGGGCTGCCAACGTCGTGCGCTCCACC 2972

RESULT 4

AL508808/c

LOCUS

DEFINITION

AL508808 Hordeum vulgare Barke developing caryopsis (3.-15.DAP)

AL508808 Hordeum vulgare cDNA clone HY09N21V 5', mRNA sequence.

AL508808.1 GI:12035311

EST.

KEYWORDS

AL508808.1 GI:12035311

EST.

SOURCE

Hordeum vulgare.

Hordeum vulgare

Hordeum vulgare

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidaeae ; Triticeae; Hordeum.

1 (bases 1 to 700)

REFERENCE

1 Michalek,W., Weschke,W., Pleissner,K.-P. and Graner,A.

AUTHORS

EST sequencing and analysis in barley

TITLE

Unpublished (2000)

JOURNAL

CONTACT: Michalek W

INSTITUTE for Plant Genetics and Crop Plant Research

CORRENSSTR.3, D-06466 Gatersleben, Germany

EMAIL: michalek@ipk-gatersleben.de, http://pgrc.ipk-gatersleben.de

Seq primer: T3 primer for 5'end.

FEATURES

source

location/Qualifiers

1..700

/organism="Hordeum vulgare"

/cultivar="Barke"

/db_xref="taxon:4513"

/clone="HY09N21V"

/clone_1b="Hordeum vulgare Barke developing caryopsis (3.-15.DAP)"

/tissue_type="developing caryopsis (3.-15.DAP)"

/lab_host="XL0LR"

/note="Vector: plasmid pBK-CMV; Site_1: EcoRI; Site_2: XhoI; mRNA was made from developing caryopsis (3.-15.DAP) of spring barley variety 'Barke', a high quality malting variety. Cloning sites: EcoRI (5'-end of cDNA) and XhoI (3'-end of cDNA). NOTE: Due to a cloning artefact caused by the k1t, in most cases the EcoRI site is NOT present, as well as the EcoRI adapter. Average insert size is 1 kb Sequence trimming: Vector sequences and sequence ends were trimmed from the 5'-and 3'-end until a 50 bp window contains less than two ambiguities. The maximum length was set to 700 bp"

BASE COUNT

152 a 204 c 244 g 93 t 7 others

ORIGIN

Alignment Scores:

pred. No.: 0.0195 length: 700

Score: 115.50 Matches: 45

Percent Similarity: 43.75% Conservative: 25

Best Local Similarity: 28.12% Mismatches: 61

Query Match: 4.80% Indels: 29

DB: 9 Gaps: 6

US-09-825-414-66 (1-487) x AL508808 (1-700)

QY 311 LeuAsnAlaGlyLysArgMetAlaGly-----LeuProLeu 322

Db 664 GTTCCCTTGGGTCTCGGGTCTCGGGCCGAGCAGCTTCTTTCTCTGTGTTCTTTC 605

QY 323 AspValAlaThrAspGlyLeuLysAlaValArgSerLeuValSerAlaThrSerLeuThr 342

Db 604 -----GTCTGGGGGCTCTCAAGCCGGTG-----CTTGCCCTCAACNGCTNCACC 563

QY 343 LysAsnGlyLeuAlaLeuAlaGlyGlyTyrAlaGlyValSerLysLeuGlnLysMetAla 362
Db 562 AACTTCGGTCTCTGT--TCCCTCGGGCAGGGGAGCTTCCGGCTCTTGGTCTCGGTCA 506
QY 363 ThrLysAsnIleThrAspSerAlaThrLysAlaAlaValSerGlnLeuSerAsnLeuVal 382
Db 505 ACTGTCGGGCTGACAGGGGCTCTCATCGGGCGGNCCGGCTGACACAACCTCTGTCTCG 446
QY 383 GlySerVal-----GlyValPheAlaGlyTyrThrThrAlaGly 395
Db 445 GCCTCCTGTGTCCTCCGCTGAGAGCTGCNMGGGCTTTTCCGACAGCTCGGTACAGCGCGC 386
QY 396 LeuAlaThrAspProAlaValLysLysAlaGluSerPheIleGlnAspLysValLysSer 415
Db 385 GCAGCACT-----TCCCTCCTTACAGGGCTCTCCGCGCAGCGACA 347
QY 416 ThrAlaSerSerThrThrSerTyrValAlaAspGlnThrValLysLeuAlaLysThrVal 435
Db 346 ACGGCCGGCTCGACAGCCTCAGCCGTTGCCACTCGACTGTGTGACAGCGCGGCTCGACA 287
QY 436 LysAspMetSerGlyGluAlaIleSerSerThrGlyAlaSerLeuArgSerThrValAsn 455
Db 286 ATCTCCACTGTGTCACCGTAGCCTCTCGACCGCAGACNACGGCCGGCTCGACGACCTCC 227

RESULT 5
CNS076CM/c 1094 bp DNA linear GSS 07-JUL-2001
LOCUS T3 end of clone BB0AA002F11 of library BB0AA from strain CBS 4732
DEFINITION of Pichia angusta, genomic survey sequence.
ACCESSION AL431228
VERSION AL431228.1 GI:12214640
KEYWORDS GSS.
SOURCE Pichia angusta.
ORGANISM Pichia angusta.
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
Saccharomycetales; Saccharomycetaceae; Pichia.

REFERENCE 1 (bases 1 to 1094)
AUTHORS Souciet,J.L., Algle,M., Artiguenave,F., Blandin,G.,
Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S.,
de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Lorente,B.,
Malpertuy,A., Neveuglise,C., Ozier-Kalogeropoulos,O., Potier,S.,
Saurin,W., Tekala,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
Wincker,P. and Weissenbach,J.
TITLE Genomic exploration of the hemiascomycetous yeasts: 1. A set of
yeast species for molecular evolution studies
JOURNAL FEBS Lett. 487 (1), 3-12 (2000)
MEDLINE 20584711
PUBMED 11152876

REFERENCE 2 (bases 1 to 1094)
AUTHORS Blandin,G., Lorente,B., Malpertuy,A., Wincker,P., Artiguenave,F.
and Dujon,B.
TITLE Genomic exploration of the hemiascomycetous yeasts: 13. Pichia
angusta
JOURNAL FEBS Lett. 487 (1), 76-81 (2000)
MEDLINE 20584723
PUBMED 11152888

REFERENCE 3 (bases 1 to 1094)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage,
2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
segref@genoscope.cns.fr - Web : www.genoscope.cns.fr)

COMMENT This GSS is part of a random genomic sequencing program of thirteen
yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
exiguus, Saccharomyces servazii, Zygosaccharomyces rouxii,
Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,
Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
5 kb were prepared and both extremities were sequenced. See
keywords for description of this sequence and for the sequence of
the other extremity of this insert.

FEATURES
Location/Qualifiers

source 1. 1094
/organism="Pichia angusta"
/strain="CBS 4732"
/db_xref="taxon:4905"
/clone="BB0AA002F11"
/clone_lib="BB0AA"
/note="end : T3"
BASE COUNT 320 a 245 c 321 g 207 t 1 others
ORIGIN
Alignment Scores: 0.0719 Length: 1094
Pred. No.: 113.50 Matches: 94
Score: 33.05% Conservative: 63
Percent Similarity: 19.79% Mismatches: 163
Best Local Similarity: 4.72% Indels: 155
Query Match: 17 Gaps: 17
US-09-825-414-66 (1-487) x CNS076CM (1-1094)
QY 5 GlnSerAlaGlnGlnProProGlyValAlaMetGluSerPheArgThrAlaSerAspAla 24
Db 1035 CAGACTTCTAGCCACAGACACCCTCTCCAGCCAAAGCTCA-----ACTTCCAAACGAGCT 982
QY 25 SerLeuAlaSerSerValArgSerValArgSerThrThrSerCysArgAspLeuGlnAla 44
Db 981 TCAGGTGATCCAGCTCTACTGCTCTGCTCA--AGTAATGTCTGTACTCTGATACC 925
QY 45 IleThrAspTyrLeuLysHisHisValPheAlaAlaHisArgPheSerValIleGlySer 64
Db 924 ATTACC-----AGATCTGCCGTCAAGTAGTGATCGGGCTCG 889
QY 65 ProAspGluArgAspAlaAlaLeuAlaHisAsnGlnGlnIleAspAlaLeuValGluThr 84
Db 888 TCCAGCGGTTCTAGCTCTGGCACAGGTTATGA----- 856
QY 85 ArgAlaAsnArgLeuTyrSerGluGlyGluThrProAlaThrIleAlaGluThrPheAla 104
Db 855 -----TCTAGCACTGCTATCTCCGATTCCGGAACCTTCAGC 820
QY 105 LysAlaGluLysPheAspArgLeuAlaThrThrAlaSerSerAlaPheGluAsnThrPro 124
Db 819 TCTGCT-----ACCTCAAGCTCAGGTTAT-----ACGTCA 790
QY 125 PheAlaAlaAlaSerValLeuGlnTyrMetGlnProAlaIleAsnLysGlyAspTyrLeu 144
Db 789 AGCGCCGCCACTCTGGATTAGTTAGGTCTACTTCAAGCGGCTCAAGGTTCCAGCTCT 730
QY 145 AlaThrProLeuLysProLeuThrProLeuIleSerGlyAlaLeuSerGlyAlaMetAsp 164
Db 729 TCCAGCTCTGCCACTCTCGATCAAGCCTTTTCAGCTCTGCTGTCTCCGGCTCCCGTGCC 670
QY 165 GlnValGlyThrLysMetMetAspArgAlaArgGlyAspLeuHisTyrLeuSerThrSer 184
Db 669 GCTACTTCTGCCCTCAAGTTATGAGTCA-----AACACAGCT 634
QY 185 ProAspLysLeuHisAspAlaMetAlaValSerValLysArgHisSerProAlaLeuGly 204
Db 633 AGCTCTGGTTTACGCTCTGCCGCTGCTATATCAAGGTTTGAGCTCTGCTTCAAGTTATGA 574
QY 205 ArgGlnValValAspMetGlyIleAlaValGlnThrPheSerAlaLeuAsnValValArg 224
Db 573 TCAAGGACT-----ACTGTCTTAGTTAAGC-----TCT 544
QY 225 ThrValLeuAlaProAlaLeuAlaSerArgProSerValGlnGly-----AlaVal 241
Db 543 GCCGCTGTGTTCCAGTTATGCTCAAGCTCTACTGTGCTGTGTTGAGCTCTGCTCA 484
QY 242 AspPheGlyValSerThrAlaGlyGlyLeuValAlaAsnAlaGlyPheGlyAspArgMet 261
Db 483 AGTTACGATTATACACCGCAGGC----- 460
QY 262 LeuSerValGlnSerArgAspGlnLeuArgGlyGlyAlaPheValLeuGlyMetLysAsp 281

[illegible]

```

source
    1. .833
       /organism="Vitis vinifera"
       /cultivar="Ugni blanc"
       /db_xref="taxon:29760"
       /clone="PT005F08"
       /clone_1lb="Veraison Grape berries SuperscriptTM Plasmid library"
       /dev_stage="Veraison stage"
       /note="Organ: Fruit; Vector: pSPORT1; site_1: Salt; Site_2: NotI; Oriented library"

BASE COUNT      183 a      251 c      137 g      262 t
ORIGIN

Alignment Scores:
Pred. NO.:          0.0757           Length:          833
Score:             111.50            Matches:         46
Percent Similarity: 45.77%           Conservative:   19
Best Local Similarity: 32.39%        Mismatches:     54
Query Match:       4.63%              Indels:         23
DB:                14                 Gaps:           6

US-09-825-414-66 (1-487) x BQ800257 (1-833)

QY      336 ValSerAlaThrSerLeuThrlLysAsnGlyLeuAlaLeuAlaGlyTyrAlaGlyVal 355
      ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      444 GTGCTGGCTTCCTCTGCCTGTCAAGCTGGTTGGCTTCTTCTTCTTGCGCCAGCT 503

QY      356 SerLysLeuGlnLysMetAlaThrLysAsnIleThrAspSerAlaThrLysAlaAlaVal 375
      ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      504 TCC----- -TCCTGCTGGGGCTGCAGCTGCCGCC 530

QY      376 SerGlnLeuSerAsnLeuValCylserValGlyValPheAlaGlyTrpThrAlaGly 395
      ||| ::||:::|| | || | || | || | || | || | || | || | || |
Db      531 TCAGTAGTCTCTTGTGGCCTTCTCAACAGCGGCTTCTGCC-----TCCACTGCTGA 584

QY      396 LeuAlaThrAspProAlaValLysLysAlaGluSerPheIleGlnAspLysValLysSer 415
      ::::: ||:::|| | || | || | || | || | || | || | || | || |
Db      585 GTCTCT-----CCTTCGGGT-----TTCTCTGCTTCTGCCTCCTCAACC 623

QY      416 ThrAlaSerSerThrThrSerTyrValAlaAspGlnThrValLysLeuAlaLysThrVal 435
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      624 CGCGCTTCCTCTACAACCTCTTGTCTCAACTTCAACTTCAACAGAGCTTCAGCTTTG 683

QY      436 LysAspMetSerGlyGluAlaIleSerSerThrGlyAlaSerLeuArgSerThrValAsn 455
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      684 GCTTCAGCTTCAGGTGCCGCTGCCTCCTCCACTGGAGCCTCT-----TCCTTGTTCCT 737

QY      456 AsnLeuArgHisArgSerAlaProGluAlaAspIleGlu---GluGlyGlyIleSerAla 474
      ::::: ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db      738 TCAGTAAGTGGCTCTGAGACACCGGTGGAGCCTCTACCTCTTTACGGCTCCCTCCACCC 797

QY      475 PheSer 476
      ::|||
Db      798 TGGTCA 803

RESULT 7
LOCUS      AL508002               680 bp      mRNA      linear      EST 04-JAN-2001
DEFINITION Hordeum vulgare Barke developing carypsops (3.-15.DAP)
VERSION    AL508002
XREFS     AL508002
KEYWORDS   Hordeum vulgare cDNA clone HY07H18V 5', mRNA sequence.
SOURCE     EST.
ORGANISM   Hordeum vulgare.
            Hordeum vulgare.
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae
            ; Triticeae; Hordeum.
            1 (bases 1 to 680)
REFERENCE  Michalek,W., Wescske,W., Pleissner,K.-P. and Graner,A.
TITLE      EST sequencing and analysis in barley
JOURNAL    Unpublished (2000)
COMMENT    Contact: Michalek W

```


OY 436 LysAspMetSerGlyLAlaIleSerSerThrGlyAlaSerLeuArgSerThrValAsn 455
Db 666 GCTTCAGCTTCAGGTGCGCGCTGCTTCACACTGAGCCTCT-----TCCTTGTCTTCT 719
OY 456 AsnLeuArgHisArg-----SerAlaProGluAla 465
Db 720 TCAGTACTGGCTCTGGAGCAGCCGGTGAGCCTCTACCTCTTTTACGGCTCCCTCTTCC 779
OY 466 AspIleGluGlyGlyGlyIleSerAlaPheSer 476
Db 780 ACCTTGTTCAATTCCTCGGTTTTCATTTCTCC 812
RESULT 9
CNS0792S
LOCUS 1150 bp DNA linear GSS 08-JUL-2001
DEFINITION T3 end of clone XBB0AA002D09 of library XBB0AA from strain CBS 4732
ACCESSION AL435950
VERSION AL435950.1 GI:12219363
KEYWORDS GSS.
SOURCE Pichia angusta.
ORGANISM Pichia angusta
REFERENCE Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
AUTHORS Saccharomycetales; Saccharomycetaceae; Pichia.
1 (bases 1 to 1150)
Souciet,J.L., Algle,M., Artiguenave,F., Blandin,G.,
Bollotin-Fukuhara,M., Bon,E., Brothier,P., Casaregola,S.,
de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B.,
Malpertuy,A., Neuvéglise,C., Ozier-Kalogeropoulos,O., Potier,S.,
Saurin,W., Tekala,F., Toffano-Nioche,C., Wesolowski-Louvel,M.,
Wincker,P. and Weissenbach,J.
Genomic exploration of the hemiascomycetous yeasts: 1. A set of
Yeast species for molecular evolution studies
JOURNAL FEBS Lett. 487 (1), 3-12 (2000)
MEDLINE 20584711
PUBMED 11152876
REFERENCE 2 (bases 1 to 1150)
AUTHORS Blandin,G., Llorente,B., Malpertuy,A., Wincker,P., Artiguenave,F.
and Dujon,B.
TITLE Genomic exploration of the hemiascomycetous yeasts: 13. Pichia
angusta
JOURNAL FEBS Lett. 487 (1), 76-81 (2000)
MEDLINE 20584723
PUBMED 11152888
REFERENCE 3 (bases 1 to 1150)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage,
2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail :
segref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT This GSS is part of a random genomic sequencing program of thirteen
yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces
exiguus, Saccharomyces servazii, Zygosaccharomyces rouxii,
Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces
lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia
angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila,
Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to
5 kb were prepared and both extremities were sequenced. See
keywords for description of this sequence and for the sequence of
the other extremity of this insert.
FEATURES
source
1..1150
/organism="Pichia angusta"
/strain="CBS 4732"
/db_xref="taxon:4905"
/clone="XBB0AA002D09"
/clone_1lb="XBB0AA"
/note="end : T3"
BASE COUNT 250 a 352 c 258 g 287 t 3 others
ORIGIN
Alignment Scores:

Pred. No.: 0.434 Length: 1150
Score: 107.00 Matches: 63
Percent Similarity: 42.70% Conservative: 57
Best Local Similarity: 22.42% Mismatches: 107
Query Match: 4.45% Indels: 54
DB: 17 Gaps: 13
US-09-825-414-66 (1-487) x CNS0792S (1-1150)
OY 216 ThrPheSerAlaLeuAsnValValArgThrValLeuAlaProAlaLeuAlaSerArgPro 235
Db 114 ACTGCTTCTAGTATTAATGAGCCAGTCGTCAGTCGTATCT-----AGCAGTACAGGG 164
OY 236 SerValGlnGlyAlaValAspPheGlyValSer-----Thr 247
Db 165 GCATCCAGTGGGGCATCCATGTTGCGCTCCAGCAGTGTCTTCGCTCTAGCCAGATTACT 224
OY 248 AlaGlyGlyLeuValAlaAsnAlaGlyPheGlyAspArgMetLeuSerValGlnSerArg 267
Db 225 GCAACGAGCAGCCGCTTCCACCTATGGCTTCAGCTCTGTTCAATATCCACATCAGGATAC 284
OY 268 AspGlnLeuArgGlyGlyAlaPheValIleuGlyMetLysAspLysGluProLysAlaAla 287
Db 285 AATTCA--AGGGCFTTAGCGCAACTGCCGA-----GCCTCAGTGCA 326
OY 288 LeuSerGluGluThrAspTrpLeuAspAlaTyrLysAlaIleLysSerAlaSerTyrSer 307
Db 327 ATTTCCAGCCAAAC-----AACGCATCC 350
OY 308 GlyAlaAlaLeuAsnAlaGlyLysArgMetAlaGlyLeuProLeuAspValAlaThrAsp 327
Db 351 AGCAGTACTTGTGCAACGGTTCA-----AGTCATATTGCTGTCTACCGCTTCTGGC 404
OY 328 GlyLeuLysAlaValAlaArgSerLeuValSerAlaThrSerLeuThrLysAsnGlyLeuAla 347
Db 405 AGTGTCCTTAGTGAAGCTAGTGAATTTCAGTGTCTGCTGGCAGCTCGAGTAATGTT 464
OY 348 LeuAlaGlyGlyTyrAlaGlyValSerLysLeuGlnLysMetAlaThrLysAsnIleThr 367
Db 465 TTCAGT-----GCTGCCCCAGCTTCATTTTCACACAGCGCTTCAAGCCAAAGTCACG 515
OY 368 AspSerAlaThrLysAlaAlaValSerGlnLeuSerAsnLeuValGlySerValGlyVal 387
Db 516 ACGGTGAGTTCATCTAGTGTGCTGCTCCTAGCGCTTCCAGTGAACCTCCAACTTCCTAAG 575
OY 388 PheAla-----GlyTrpThrThrAlaGlyLeuAlaThrAspProAlaValLysLysAla 405
Db 576 ACTGCTTCGAGCGGCTTCTCCAGCTCCCTCAGCGCAACGTC-----CGCGGTGCG 626
OY 406 GluSerPheIleGlnAspLys-----ValLysSerThrAlaSerSerThr 420
Db 627 GACAGCTTTCACAGCAGCACCACGCGTCTATTGTCTCGAGCAACACTGCTTCCAGTGT 686
OY 421 ThrSerTyrValAlaAspGlnThrValLysLeuAlaLysThrValLysAspMetSerGly 440
Db 687 TCTAGCTATGTTTCCGACCAGGCC-----AGCGTCACATCTAGCTCAGGT 731
OY 441 GluAlaIleSerSerThrGlyAlaSerLeuArgSer---ThrValAsnAsnLeuArgHis 459
Db 732 GGCTCTAGCGAAATATCACAGTTTATCTTGCTAGTGCCACCATCAGCTCTGTGTCAAACA 791
OY 460 ArgSerAlaProGluAlaAspIleGluGlyGlyIleSerAlaPheSerArgSerGlu 479
Db 792 GCAATTCAGGCCAAAGTAGTCATCCAAATGCTGAGTCAGC-----TCCGAA 839
OY 480 Thr 480
Db 840 ACC 842
RESULT 10
LOCUS AY107053 1763 bp mRNA linear HTC 25-MAY-2002
DEFINITION Zea mays PCC082530 mRNA sequence.

ACCESSION	AY107053
VERSION	AY107053.1
KEYWORDS	GI:21210131
SOURCE	HTC.
ORGANISM	Zea mays.
	Zea mays
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACC clade; Panicoideae; Andropogoneae; Zea.
AUTHORS	1 (bases 1 to 1763) Hainey,C.F., Dolan,M., Miao,G.H., Vogel,J.M., Whitsitt,M.S., Arthur,L.W., Hanafey,M., Morgante,M. and Tingey,S.V.
TITLE	Maize Mapping Project/Dupont Consensus Sequences for Design of Overgo Probes
JOURNAL	Unpublished (2002)
REFERENCE	2 (bases 1 to 1763)
AUTHORS	Coe,E.C.
TITLE	Direct Submission
JOURNAL	Submitted (25-APR-2002) Maize Mapping Project, University of Missouri, Columbia, MO 65211, USA
FEATURES	location/Qualifiers
Source	1..1763 /organism="Zea mays" /db_xref="MaizeDB:634619" /db_xref="taxon:4577" /clone="PC0082530" /clone_lib="Maize Mapping Project/Dupont Consensus Library" /note="this sequence is part of a project of EST assemblies resulting from the application of public contigs to seed Dupont contigs; this resource was assembled by Dupont as part of a collaboration for the overgo addressing of BACs in conjunction with the Maize Mapping Project"
BASE COUNT	341 a 543 c 535 g 344 t
ORIGIN	
Alignment Scores:	
Pred. No.:	0.913 Length: 1763
Score:	107.00 Matches: 127
Percent Similarity:	34.14% Conservative: 56
Best Local Similarity:	23.69% Mismatches: 189
Query Match:	4.45% Indels: 166
DB:	11 Gaps: 25
US-09-825-414-66 (1-487) x AY107053 (1-1763)	
QY 19	ArgThrAlaSerAspAlaSerLeuAlaSerSerValArgSerValSerThr--Thr 37
Db 217	::
QY 38	SerCysArgAspLeuGlnAlaIleThrAspTyrLeuLysHisValPheAlaAlaHis 57
Db 277	::
QY 58	ArgPheSerValIleGlySerProAspGluArgAspAlaAlaLeuAlaHisAsnGluIn 77
Db 295	-----CCAAAGCTCCAA-----
QY 78	IleAspAlaLeuValGluThrArgAlaAsnArgLeuTyrSerGluGlyGluThrProAla 97
Db 307	-----GCAACGACCGCGACCGCTTCGTTCTCCGGTGGAGGCCCTGCGCC 351
QY 98	ThrIleAlaGluThrPheAlaLysAlaGluLysPheAspArgLeuAlaThr----- 114
Db 352	GCCTCAGCGCACCTCTGGCCCGCTATTGCGCGCGCGCCGCCACGCGCTCTCGCCGTCA 411
QY 115	-----ThrAlaSerSerAlaPheGluAsnThrProPheAla 126
Db 412	TTACAGACGTGACCCCTGACCTCCACAGTCATCCCATTCGCCAAGAGAGCTCGCGTTACGT 471
QY 127	AlaAlaSerValLeuGlnTyrMetGlnProAlaIleAsnLysGlyAspTyrPleuAlaThr 146
Db 472	GCCACGTCCTCTTCCCTCCAGCGCCACCATGCTCTCGCTCC-----TTGCTTACA 522

[illegible]

QY 450 euArgSerThrValAsnAsnLeuArgHisArgSerAlaProGluAlaAspIleGluGluG 470
|||||
Db 1433 ACCGAGCAGGAGGAGCTTTATTTCACGCCTCAACCAACCGTGGCTTGATTCGCTC 1492
QY 470 IyGlyIleSerAlaPheSerArgSerGluThrProPheGlnLeu 484
|||||
Db 1493 GTGGGGTTCCTGATTTCCACAAGTTGTGTCCCTTACTCCCTA 1536
RESULT 11
BQ800379
LOCUS BQ800379 841 bp mRNA linear EST 30-JUL-2002
DEFINITION EST 7414 Veralson Grape berries SuperScriptTM Plasmid library Vitis
vinfifera cDNA clone PT009E05 3', mRNA sequence.
ACCESSION BQ800379
VERSION BQ800379.1 GI:22015345
KEYWORDS EST.
SOURCE Vitis vinfifera.
ORGANISM Vitis vinfifera
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Vitaceae; Vitis.
REFERENCE 1 (bases 1 to 841)
AUTHORS Abbal,P., Agasse,A., Ageorges,A., Atanassova,R., Barrieu,F.,
Couture,C., Dedaldechamp,F., Delrot,S., Glissant,D., Grimplet,J.,
Hamdi,S., Romieu,C. and Terrier,N.
TITLE Generation of Expressed Sequence Tag from Grape Berry (skin, pulp
or seeds) at Various Developmental Stages
JOURNAL Unpublished (2002)
COMMENT Contact: Delrot S.
TRANSPORT DES ASSIMILATS
UMR CNRS 6161, Universite de Poitiers
Batiment Botanique, 40 Avenue du Recteur Pineau, 86022 Poitiers
Cedex, France
Tel: 00-33-(0)5-49-45-41-85
Fax: 00-33-(0)5-49-45-41-86
Email: Serge.Delrot@univ-poitiers.fr
Seq primer: SP6.
FEATURES
source Location/Qualifiers
1..841
/organism="Vitis vinfifera"
/cultivar="Ugni blanc"
/db_xref="taxon:29760"
/clone="PT009E05"
/clone_lib="Veralson Grape berries SuperScriptTM Plasmid
library"
/dev_stage="Veralson stage"
/note="Organ: Fruit; Vector: pSPORT1; Site_1: SalI;
Site_2: NotI; Oriented library"
BASE COUNT 187 a 251 c 138 g 265 t
ORIGIN
Alignment Scores:
Pred. No.: 0.373 Length: 841
Score: 105.50 Matches: 48
Percent Similarity: 44.08% Conservative: 19
Best Local Similarity: 31.58% Mismatches: 56
Query Match: 4.38% Indels: 29
DB: 14 Gaps: 7
US-09-825-414-66 (1-487) x BQ800379 (1-841)
QY 336 ValSerAlaThrSerLeuThrLysAsnGlyLeuAlaLeuAlaGlyGlyTyraAlaGlyVal 355
|||||
Db 444 GTGCTGGCTTCCTCTGCTGCTTCACGCTGGTTGGCTTCTTCTCTCTGCGGCGAGCT 503
QY 356 SerLysLeuGlnLysMetalThrLysAsnIleThrAspSerAlaThrLysAlaAlaVal 375
|||||
Db 504 TCC-----TCTGCTGGGGCTGCAGCTGCCGCC 530
QY 376 SerGlnLeuSerAsnLeuValGlySerValGlyValPheAlaGlyTrpThrThrAlaGly 395
|||||
Db 531 TCAGTAGTCTCTCTTCTTGGCCTTCTCAACAGCGGCTTCTGCC-----TCCACTGCTGGA 584

QY 396 LeuAlaThrAspProAlaValLysLysAlaGluSerPheIleGlnAspLysValLysSer 415
:::::
Db 585 GTCTCT-----CCTTCGGTT-----TTCTCTGCTTCTGCTCCTCAACC 623
QY 416 ThrAlaSerSerThrThrSerTyrrValAlaAspGlnThrValLysLeuAlaLysThrVal 435
|||||
Db 624 GCGGCTTCCTTACACCTCCTTGTCTCACTTCACACTTCACACAGAGCTTCAGCTTGG 683
QY 436 LysAspMetSerGlyGluAlaIleSerSerThrGlyAlaSerLeuArgSerThrValAsn 455
|||||
Db 684 GCTTCAGCTTCAGGTGCGCGCTGCCTCCTCCACTGCAGCCTCT-----TCCTTGTTCCT 737
QY 456 AsnLeuArgHisArgSerAlaProGluAlaAspIleGluGluGlyGlyIleSer--Ala 474
:::::
Db 738 TCAGTAAGTGGCTGTGAGCA-----GCCGGTGAGCCTTACCTCT 779
QY 475 PheSerArgSerGluThrProPheGlnLeuArgArg 486
|||||
Db 780 TTTACGGCTCCTCTTCACACCTTGTTCATTCCTCGG 815
RESULT 12
BF265543/c
LOCUS BF265543 850 bp mRNA linear EST 23-OCT-2001
DEFINITION HV_Cea0012J12f Hordeum vulgare seedling green leaf EST library
HVCDA0004 (Blumeria challenged) Hordeum vulgare cDNA clone
HV_Cea0012J12f, mRNA sequence.
ACCESSION BF265543
VERSION BF265543.2 GI:13262162
KEYWORDS EST.
SOURCE Hordeum vulgare.
ORGANISM Hordeum vulgare.
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae
; Triticeae; Hordeum.
REFERENCE 1 (bases 1 to 850)
AUTHORS Wing,R., Close,T.J., Kleinhofs,A., Wise,R., Wel,F., Begum,D.,
Frisch,D., Yu,Y., Henry,D., Palmer,M., Rambo,T., Simmons,J., Choi
D.W., Fenton,R.D., Oates,R. and Main,D.
TITLE Development of a genetically and physically anchored EST resource
for barley genomics: Blumeria infected incompatible (Mla13)
seedling leaf cDNA library
JOURNAL Unpublished (2001)
COMMENT On Nov 17, 2000 this sequence version replaced gi:11196537.
Contact: Wing RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Total hg bases = 429
Seq primer: AATTAAACCTCACAATAAGCG
High quality sequence stop: 743.
FEATURES
source Location/Qualifiers
1..850
/organism="Hordeum vulgare"
/cultivar="C116155 (Mla13)"
/db_xref="taxon:4513"
/clone="HV_Cea0012J12f"
/clone_lib="Hordeum vulgare seedling green leaf EST
library HVCDA0004 (Blumeria challenged)"
/issue_type="seedling green leaf"
/lab_host="TJC121"
/note="Vector: lambdaZAP; Site_1: EcoRI; Site_2: XhoI;
C.I. 16155 (Mla13) plants were greenhouse grown in the R
wise lab at Iowa State University, Ames, IA; 7 day old
green seedlings were challenged with isolate A27 (AvrMla13
) of Blumeria graminis f. sp. hordei, and leaves were
harvested 20 and 24 hr post-inoculation and snap frozen;
uninoculated leaves were harvested 20 hr post-inoculation
(wel, wise). In the TJ Close lab at the University of
California, Riverside, total RNA was prepared from each
sample pool, equal quantities of all three RNA pools were

combined, poly(A) RNA was purified from the mixture, one cDNA library was made, and 1 million pfu were in vivo excised to give Bluescript SK(-) cDNA phagemids (Choi, Close). Phagemids were plated and picked at the Clemson University Genomics Institute (CUGI) (Begum, Palmer, Frisch, Atkins and Wing). Plasmid DNA preparations, DNA sequencing and sequence analysis were performed at CUGI (Wing, Yu, Frisch, Henry, Simmons, Oates, Rambo, Main). The sequence has been trimmed to remove vector sequence and contains a minimum of 100 bases of phred value 20 or above. For more details on library preparation and sequence analysis see

http://www.genome.clemson.edu/projects/barley. To order this clone see http://www.genome.clemson.edu/orders Also see Close TJ, Wing R, Kleinofs A, Wise R (2001) Genetically and physically anchored EST resources for barley genomics. Barley Genetics Newsletter 31:29-30. (http://wheat.pw.usda.gov/ggpages/bgn/31/cover.html)"

BASE COUNT 139 a 286 c 294 g 131 t
ORIGIN

Alignment Scores:

Pred. No.: 0.38 Length: 850
Score: 105.50 Matches: 59
Percent Similarity: 36.47% Conservative: 34
Best Local Similarity: 23.14% Mismatches: 100
Query Match: 4.38% Indels: 62
DB: 12 Gaps: 9

US-09-825-414-66 (1-487) x BF265543 (1-850)

QY 11 ProGlyValAlaMetGluSerPheArgThrAlaSerAspAlaSerLeuAlaSerSerSer 30
Db 670 CCGGGGTCGCCCGCGAGCCC-----AGTGTGTACAAAGCGCCACCGCGGTAGATGAT 617
QY 31 ValArgSerValSerThrThrSerCysArgAspLeuGlnAlaIleThrAspTyrLeuLys 50
Db 616 GCTCAGGCGCTCACCAAGGGTCCGCCCAACAACGAGTAGCCCTCGATCAAGCCCATGAG 557
QY 51 HisHisValPheAlaAlaHisArgPheSerValIleGlySerProAspGluArgAspAla 70
Db 556 CACCAC-----CTGGCATGCCCGAGATGGCGAGGATGCT 524
QY 71 AlaLeuAlaHisAsnGluGlnIleAspAlaLeuValGluThrArgAlaAsnArgLeuTyr 90
Db 523 CTGCGCGGTGCACCAAGTTGGGGTTCGCCAGTAGTCGAGGCCGCTCGGAGAGAATCTG 464
QY 91 SerGluGlyGluThrProAlaThrIleAlaGluThrPheAlaLysAlaGluLysPheAsp 110
Db 463 CGCGCGGCGCTTGAACACAGACGCTCGCGGAACCTTGATGCCGTTCTTGACAG----- 410
QY 111 ArgLeuAlaThrThrAlaSerSerAlaPheGluAsnThrProPheAlaAlaAlaSerVal 130
Db 409 GATCTCGGGGAACACGACCCGAGGGCTCCGAGCATGCCCA-----CCGCGAGTG 359
QY 131 LeuGlnTyrMetGlnProAlaIleAsnLysGlyAspTrpLeuAlaThrProLeuLysPro 150
Db 358 GATCACCTCCAGCTCCCTGTTCTTGCGCAAGGTCTCTGGGTACGCCGCAATCCCGCGGT 299
QY 151 LeuThrProLeuIleSerGlyAlaLeuSerGlyAlaMetAspGlnValGlyThrLysMet 170
Db 298 GTCCACCCCTAGTGTGCCGGGGAATCTCGCGGT-----CAGGTACGACGGGGTCTG 248
QY 171 MetAspArgAlaArgGlyAspLeuHisTyrLeuSerThrSerProAspLysLeuHisAsp 190
Db 247 CTCGGAGAACGGGCCACAGGTACTTGGGACGGTCAAGGCCGTACACGAT----- 200
QY 191 AlaMetAlaValSerValLysArgHisSerProAlaLeuGlyArgGlnValValAspMet 210
Db 199 -----GCTCTGGGCGC----- 188
QY 211 GlyIleAlaValGlnThrPheSerAlaLeuAsn-ValValArgThrValLeu----- 227

Db 187 -----GCTCTTGACGGTCCGGCCGACGGTGATGCGGCCCC 152
QY 228 ----AlProAlaLeuAlaSerArgProSerValGlnGlyAlaVal----- 241
Db 151 GAAGTCGCCGACGCTGCGGACGAGGTCATCGCGCGGGCGGTGCCGAGGAAGCTGCT 92
QY 242 -----AspPheGlyValSerThrAlaGlyGly 250
Db 91 CGTCTGTGTGACGCGCCGACGCGGCCCAATTTCGCTGTGTGCTGG 49
RESULT 13
CNS076CP/c
LOCUS
DEFINITION T7 end of clone BB0AA002601 of library BB0AA from strain CBS 4732 of Pichia angusta, genomic survey sequence.
ACCESSION AL431231
VERSION AL431231.1 GI:12214643
KEYWORDS
SOURCE
ORGANISM
Pichia angusta.
Pichia angusta
Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes; Saccharomycetales; Saccharomycetaceae; Pichia.
REFERENCE
AUTHORS Souciet,J.L., Aigle,M., Artiguenave,F., Blandin,G., Bolotin-Fukuhara,M., Bon,E., Brottier,P., Casaregola,S., de-Montigny,J., Dujon,B., Durrens,P., Lepingle,A., Llorente,B., Malpertuy,A., Neuvéglise,C., Ozier-Kalogeropoulos,O., Potier,S., Saurin,W., Tekai,F., Toffano-Nioche,C., Wesolowski-Louvel,M., Wincker,P. and Weissenbach,J.
TITLE Genomic exploration of the hemiascomycetous yeasts: 1. A set of yeast species for molecular evolution studies
REFERENCE
AUTHORS FEBS Lett. 487 (1), 3-12 (2000)
TITLE
JOURNAL MEDLINE 20584711
PUBMED 11152876
REFERENCE
AUTHORS 2 (bases 1 to 888)
Blandin,G., Llorente,B., Malpertuy,A., Wincker,P., Artiguenave,F. and Dujon,B.
TITLE Genomic exploration of the hemiascomycetous yeasts: 13. Pichia angusta
JOURNAL FEBS Lett. 487 (1), 76-81 (2000)
MEDLINE 20584723
PUBMED 11152888
REFERENCE
AUTHORS 3 (bases 1 to 888)
Genoscope.
TITLE Direct Submission
JOURNAL Submitted (08-SEP-2000) Genoscope - Centre National de Sequencage, 2 rue Gaston Cremieux, CP 5706, 91057 EVRY cedex, FRANCE. (E-mail : seque@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT This GSS is part of a random genomic sequencing program of thirteen yeast species: Saccharomyces bayanus var. uvarum, Saccharomyces exiguus, Saccharomyces servazii, Zygosaccharomyces rouxi, Saccharomyces kluyveri, Kluyveromyces thermotolerans, Kluyveromyces lactis var. lactis, Kluyveromyces marxianus var. marxianus, Pichia angusta, Debaryomyces hansenii var. hansenii, Pichia sorbitophila, Candida tropicalis and Yarrowia lipolytica. Genomic inserts of 3 to 5 kb were prepared and both extremities were sequenced. See keywords for description of this sequence and for the sequence of the other extremity of this insert.
FEATURES
source
1. 888
/organism="Pichia angusta"
/strain="CBS 4732"
/db_xref="taxon:4905"
/clone="BB0AA002601"
/clone_lib="BB0AA"
/note="end : T7"
BASE COUNT 225 a 190 c 273 g 194 t 6 others
ORIGIN
Alignment Scores: 0.41 Length: 888
Pred. No.: 105.50 Matches: 71
Percent Similarity: 36.36% Conservative: 41

Best Local Similarity: 23.05%		Mismatches: 119	
Query Match: 4.38%		Indels: 77	
DB: 17		Gaps: 14	
US-09-825-414-66 (1-487) x CNS076CP (1-888)			
QY	214	ValGlnThrPheSerAlaLeuAsnValAlaArgThrValLeu-----AlaProAlaLeu	231
Db	882	CTACAGCAKTCCTCGAATGCCAACTCTAGCCAGACGATTCATCATCTGCCCTMCCCT	823
QY	232	AlaSerArgProSerValGlnGlyAlaValaAlaAspPheGlyValSerThrAlaGlyLeu	251
Db	822	GCTTCA---TCGTCGACAAAGCTCATTTGGCGAAGCCTTTTCGAGTGTGGAAGCACT	766
QY	252	ValAlaAsnAlaGlyPheGlyAspArgMetLeuSerVal---GlnSerArgAspGlnLeu	270
Db	765	TCTGCCAGAGCC-----TCCCGTACGGCGGAATATCTCAATCAGCCAGCCAGCAT	715
QY	271	ArgGlyGlyAlaPheValLeuGlyMetLysAspLysGluProLysAlaAlaLeuSerGlu	290
Db	714	GGCTCCACTGCTTCTAGTATTAATGSCAG-----TCSTCAGTCGTATCTAGC	667
QY	291	GluThr-----AspTrpLeuAspAlaTyrLysAlaIleLysSerAlaSerTyrSer	307
Db	666	AGTACAGGGGCATCCAGTTGGGCATCCATGTTCGGCTCC--AGCAGTGTCTTCGCTCT	610
QY	308	Gly-----AlaAlaLeuAsnAlaGlyLysArgMetAlaGlyLeuProLeuAspValAla	325
Db	609	AGCCAGATTACTGCAACGACGACCCTTCACCCTATGGCTTCAGCTGTTCATATATCC	550
QY	326	ThrAspGlyLeu-----LysAlaValaArg	333
Db	549	ACATCAGGATACAAATTCACAGGGCTTTAGCCCACTGCCGAGCGTCAAGTGCATTTCC	490
QY	334	SerLeuValSerAlaThrSerLeuThr---LysAsnGlyLeu-----	346
Db	489	AGCCAAACACAGCATCCAGCAGTACTTTGTGCAACGGTTCAAGCTCAATTGCTGCTCT	430
QY	347	-----AlaLeuAlaGlyGlyTyrAlaGlyValSerLysLeuGlnLysMetAla	362
Db	429	ACCGCTTCTGGCAGTGTCCCTAGTGGAGCTAGTGGAATTTCCAGTGTTCGTGTCGACGC	370
QY	363	ThrLysAsnIleThrAspSerAlaThrLysAlaAlaValSerGlnLeuSerAsnLeuVal	382
Db	369	TCGAGTAAATKTTTTCAGTGTGCCCCCAGCTCCATTTTCACACAGCGCTTCAAGCCAGTC	310
QY	383	GlySerValGlyValPheAlaGlyTyrTrpThrThrAlaGlyLeuAlaThrAspProAlaVal	402
Db	309	ACGACGGTGAGTTCATCTAGTGTGCTGCTCTAGCGCTTCAGTGCACACTCCACATTTCT	250
QY	403	LysLysAlaGluSerPheIleGlnAspLysVal-----	413
Db	249	AAGACTGCTTCGAGCGGCTTCTCCAGCTCCCTCAGCGCAAGCTCTCGCCGTGGACAGC	190
QY	414	-----LysSerThrAlaSerSerThrThrSer	422
Db	189	TTCTCCAGCAGCACCAAGGCTCTATTGTCTCGAGCAACACTGCTTCAGTGTTCCTAGC	130
QY	423	TyrValAlaAspGlnThrValLysLeuAlaLysThrValLysAspMetSerGlyAlaAla	442
Db	129	TATGTTTCCGACACAGGCC-----ACGCTCACATCTAGCTCAGGTGGCTCT	85
QY	443	IleSerSerThrGlyAlaSerLeuArgSerThrValAsnAsnLeuArgHisArgSerAla	462
Db	84	ACGGAATAATACCAAGTTATCCTG-----	61
QY	463	ProGluAlaAspIleGluGly	470
Db	60	CCTAGTCCACCATCAGCTCTGGT	37
RESULT 14			
BQ795937/c			
LOCUS	BQ795937	579 bp	mRNA linear EST 30-JUL-2002

DEFINITION	EST 4875 Ripening Grape berries Lambda Zap II Library Vitis		
ACCESSION	vlnifera cDNA clone RT022G09 3', mRNA sequence.		
VERSION	BQ795937		
KEYWORDS	BQ795937.1 GI:22010903		
SOURCE	EST.		
ORGANISM	Vitis vinifera.		
REFERENCE	Vitis vinifera		
AUTHORS	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Vitaceae; Vitis.		
TITLE	1 (bases 1 to 579)		
JOURNAL	Abbal,P., Agasse,A., Ageorges,A., Atanassova,R., Barrieu,F., Couture,C., Dedaldechamp,F., Delrot,S., Glissant,D., Grimplet,J., Hamdi,S., Romieu,C. and Terrier,N.		
COMMENT	Generation of Expressed Sequence Tag from Grape Berry (skin, pulp or seeds) at Various Developmental Stages		
	Unpublished (2002)		
	Contact: Romieu C.		
	Unite de Recherche des Produits de la Vigne		
	Institut National de la Recherche Agronomique		
	2, place Viala, 34 060 Montpellier Cedex 01, France		
	Tel: 00-33-(0)4-99-61-28-62		
	Fax: 00-33-(0)4-99-61-28-57		
	Email: romieu@ensam.inra.fr		
	Seq primer: T7.		
FEATURES	Location/Qualifiers		
source	1..579		
	/organism="Vitis vinifera"		
	/cultivar="Shiraz"		
	/db_xref="taxon:29760"		
	/clone="RT022G09"		
	/clone_lib="Ripening Grape berries Lambda Zap II Library"		
	/dev_stage="ripening stage"		
	/note="Organ: Fruit; Vector: Lambda Zap II; Site_1: Eco RI		
	; Site_2: XhoI; Oriented library, construction described		
	in Generation of ESTs from grape Berry (skin, pulp or		
	seeds) at various developmental stages by Terrier,N.,		
	Ageorges,A., Abbal,P., Romieu,C. In J. Plant Physiol. 158		
	(12): 1575-83 2001"		
BASE COUNT	177 a	136 c	174 g 92 t
ORIGIN			
Alignment Scores:			
Pred. No.:	0.222	Length: 579	
Score:	105.00	Matches: 45	
Percent Similarity:	43.71%	Conservative: 21	
Best Local Similarity:	29.80%	Mismatches: 53	
Query Match:	4.36%	Indels: 32	
DB:	14	Gaps: 6	
US-09-825-414-66 (1-487) x BQ795937 (1-579)			
QY	336	ValSerAlaThrSerLeuThrLysAsnGlyLeuAlaLeuAlaGlyTyrAlaGlyVal	355
Db	528	GTGCTGGCTTCTCTGCTGCTTCAGCTGTGCTTCTTCTTCTTCTTGGCGGCACT	469
QY	356	SerLysLeuGlnLysMetAlaThrLysAsnIleThrAspSerAlaThrLysAlaAlaVal	375
Db	468	TCC-----TCTCTGGGGCTGCAGCTGCTCC	442
QY	376	SerGlnLeuSerAsnLeuValGlySerValGlyValPheAlaGlyTyrThrAlaGly	395
Db	441	TCAGTAGTCTCTTGTGGCCCTTCTCAACAGCGGCTTGTCC-----TCCACTGTGGA	388
QY	396	LeuAlaThrAspProAlaValLysLysAlaGluSerPheIleGlnAspLysValLysSer	415
Db	387	GTCCTCT-----CCTTCGGTT-----TTCCTGCTTCTGCTCTCTCAACC	349
QY	416	ThrAlaSerSerThrThrSerTyrValAlaAspGlnThrValLysLeuAlaLysThrVal	435
Db	348	GGGCTTCTCTACCACTCTTGTGCTCAACTTCAACTTCAACAGAGCTTCAGCTTTG	289
QY	436	LysAspMetSerGlyLysAlaIleSerSerThrGlyAlaSerLeuArgSerThrValAsn	455

Db 288 GCTTCAGCTTCAGGTGCCCGCTGCCCTCCACTGAGACCTCT-----TCCCTTTGTTCT 235
QY 456 AsnLeuArgHisArg-----SerAlaProGluAla 465
Db 234 TCAGTAACTGGCTTCGAGACGACCGGTGAGACCTCTACTCTTTTACGGCTCCCTCTTCC 175
QY 466 AspIleGluGlyGlyIleSerAlaPheSer 476
Db 174 ACCTTGTTCAATTCCTCGGTTTTCATTCCTCC 142
RESULT 15
BM077210
LOCUS BM077210 806 bp mRNA linear EST 05-FEB-2002
DEFINITION TREST-A4136 TREST-A Hypocrea jecorina cDNA clone Tr-A4136 5', mRNA
sequence.
ACCESSION BM077210 GI:18498392
VERSION BM077210
KEYWORDS EST.
SOURCE Hypocrea jecorina.
ORGANISM Hypocrea jecorina.
REFERENCE 1 (bases 1 to 806)
AUTHORS Chambergo,F.S., Bonaccorsi,E.D., Ferreira,A.J.S., Ramos,A.S.P.,
Ferreira,J.R.Jr., Abrahamo-Neto,J., Farah,J.P.S. and El-Dorriy,H.
TITLE Elucidation of the metabolic fate of glucose in the filamentous
fungus Trichoderma reesei using expressed sequence tag (EST)
JOURNAL J. Biol. Chem. 277 (16), 13983-13988 (2002)
MEDLINE 21950703
COMMENT Contact: El-Dorriy, Hamza
Department of Biochemistry
Institute of Chemistry. University of Sao Paulo
Avenida Professor Lineu Prestes, 748, Sao Paulo, SP, 05508-900,
BRASIL
Tel: (55) 11-38183848
Fax: (55) 11-38183848
Email: dorriy@iq.usp.br
PCR Primers
FORWARD: Universal M13 forward primer
BACKWARD: Universal M13 reverse primer
Plate: 44 row: A column: 8
Seq primer: M13 reverse primer
High quality sequence stop: 806
POLYA=No.
FEATURES
source Location/Qualifiers
1..806
/organism="Hypocrea jecorina"
/strain="QM9414 (ATCC26921)"
/db_xref="taxon:51453"
/clone="Tr-A4136"
/clone_lib="TREST-A"
/sex="Asexual"
/tissue_type="Mycelia"
/dev_stage="18 hr Glycerol-grown culture"
/lab_host="E. coli SOLR cells (kanamycin resistant)"
/note="Vector: pBluescript SK(+); Site_1: EcoRI; Site_2:
XhoI; anamorph=Trichoderma reesei; Cloned unidirectionally
, 5' end of the cDNA cloned into EcoRI site of
pBluescript. Primer: Oligo (dT). Average insert size: 1,2
kb; Uni-ZAP XR Vector system -5', adaptor sequence:
5'GAATTCCGACGAG3' -3' adaptor sequence:
5'CTCGAGTTTTCCTTTTTCCTTTT3'."

Alignment Scores:
Pred. No.: 0.395 Length: 806
Score: 105.00 Matches: 61
Percent Similarity: 37.73% Conservative: 22
Best Local Similarity: 27.73% Mismatches: 70
Query Match: 4.36% Indels: 67

DB: 13 Gaps: 10
US-09-825-414-66 (1-487) x BM077210 (1-806)
QY 301 IleYsSerAlaSerTySerGlyAlaAlaLeuAsnAlaGlyLys----- 315
Db 171 TTGAAAAACGTTGGCTCGTCGCGCCAGCGCAACGTGCGCAAGGTGCTCCCGCC 230
QY 316 -----ArgMetaAlaGlyLeuProLeuAsp 323
Db 231 CTCCTCGCCGCCAACAAATTCACCGTCACCGTCCCGCGCCCTCTCCCATCC 290
QY 324 ValAlaThrAspGlyLeuLysAlaValArg---SerLeuValSerAlaThrSerLeuThr 342
Db 291 ACGTACCCGAGCGCGCTCCGCGCTCGACGTCGACTTCAGCTCCGTCGAGTCCCGACG 350
QY 343 LysAsnGlyLeuAlaLeuAlaGlyGlyTyTrAlaGlyValSerLysLeuGlnLysMeta 362
Db 351 -----GCCCGCTTGCCCGGCCAACAGACCGCTGTCACCGTCGCGCAGCCGCC 401
QY 363 ThrLysAsn-----IleThrAspSerAlaThrLysAlaAlaValSerGlnLeu 378
Db 402 CTCAAGACGAGCAAAAGAGGCTCATTTAGCGCCGCGCGCGCGCTCAAGCGCTTC 461
QY 379 -----SerAsnLeuValGlySerValGlyValPhe 388
Db 462 CTGCCCTCCGAGTTGGGCTGCGACCTGACCAAGAGCTTCCGCCCAAGCTGCCGCTTC 521
QY 389 AlaGlyTrpThrThrAlaGlyLeuAlaThrAspProAlaValLysLysAlaGlu----- 406
Db 522 GCA-----ACCAAGTTCGAGATCAGC 542
QY 407 SerPheIleGlnAspLysValLysSerThr-AlaSerSerThrThrSerTyValAla-- 425
Db 543 CGGTACCTCGAGACAGCAAGGCCAAGACGACGCGCTCTACGTACACGCGTCTACAGCGG 602
QY 426 -----AspGlnThrValLysLeuAlaLysThrValLysAspMetSerG 440
Db 603 CCCTTCTCGACTGGGACTACAGTACACTTCATCTTCAAGAGTGTGATCTAAGC-- 660
QY 440 yGluAlaIleSerSerThrGlyAlaSerLeuArgSerThrValAsnAsnLeuArgHisAr 460
Db 661 ---CCGTGCTGACGACGCGGGGCAACACCGTCTTTAACAACTTCGAACCTTCGAAGCCG 716
QY 460 gSerAla-----ProGluAlaAspIleGluGlyGly 471
Db 717 TTCCCCAGCGCGGTGCGGGTCTTTTAAACCCGAGGAAGACCAAGAAACGGGCGGT 774

Search completed: January 31, 2003, 07:16:29
Job time : 1589.62 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: January 31, 2003, 04:12:39 : Search time 54.556 Seconds
(without alignments)
2737.582 Million cell updates/sec

Title: US-09-825-414-66
Perfect score: 2407
Sequence: 1 MHINSAQQPVGVMESFRT.....EEGISAFSRSETPQLRL 487

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+pn.model -DEV=xlp
-Q/cgn2_1/USPTO_spool/US09825414/runat.28012003_161742_26114/app_query.fasta.1.1294
-DB=Issued_Patents_NA -QFMT=fastap -SUFFIX=p2n.rn1 -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=biosum62 -TRANS=human40.cdl
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09825414.@CGN_1_1_40_@runat.28012003_161742_26114 -NCPU=6 -ICPU=3
-NO_XLPXY -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG -DEV_TIMEOUT=120
-WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/lna/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/lna/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/lna/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/lna/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/lna/PCFUS_COMB.seq:*
6: /cgn2_6/ptodata/1/lna/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	146.5	6.1	9542	4 US-08-968-685A-9	Sequence 9, Appl1
2	140	5.8	7100	4 US-09-308-375-1	Sequence 1, Appl1
3	132	5.5	3300	1 US-08-194-290-6	Sequence 6, Appl1
4	132	5.5	3300	2 US-08-614-377A-6	Sequence 6, Appl1
5	132	5.5	3300	4 US-09-142-648B-6	Sequence 6, Appl1
6	131.5	5.5	5393	2 US-08-591-079-9	Sequence 9, Appl1
7	130	5.4	46819	4 US-09-453-702B-72	Sequence 72, Appl1
8	127.5	5.3	4403765	4 US-09-103-840A-2	Sequence 2, Appl1
9	126.5	5.3	32768	4 US-08-961-527-71	Sequence 71, Appl1
10	126.5	5.3	4411529	4 US-09-103-840A-1	Sequence 1, Appl1
11	125.5	5.2	3666	2 US-08-682-517-13	Sequence 13, Appl1
12	125.5	5.2	3666	2 US-08-682-517-14	Sequence 14, Appl1

13	125.5	5.2	4197	2 US-08-682-517-7	Sequence 7, Appl1
14	125.5	5.2	4197	2 US-08-682-517-8	Sequence 8, Appl1
15	123	5.1	2489	3 US-09-141-047-7	Sequence 7, Appl1
16	122.5	5.1	3123	4 US-09-134-001C-2528	Sequence 2528, Ap
17	122	5.1	3979	4 US-09-085-199B-10	Sequence 10, Appl
18	122	5.1	17612	3 US-08-911-853-29	Sequence 29, Appl
19	122	5.1	17612	4 US-09-479-409-29	Sequence 29, Appl
20	122	5.1	17612	4 US-09-479-453-29	Sequence 29, Appl
21	121	5.0	2550	4 US-08-961-527-364	Sequence 364, App
22	120.5	5.0	4411529	4 US-09-103-840A-1	Sequence 1, Appl1
23	119.5	5.0	4483	4 US-08-961-527-363	Sequence 363, App
24	119	4.9	25165	4 US-09-453-702B-39	Sequence 39, Appl
25	117.5	4.9	1464	4 US-08-809-326A-3	Sequence 3, Appl1
26	117.5	4.9	1947	4 US-08-809-326A-17	Sequence 17, Appl
27	117	4.9	2742	3 US-08-911-853-16	Sequence 16, Appl
28	117	4.9	2742	4 US-09-479-409-16	Sequence 16, Appl
29	117	4.9	2742	4 US-09-479-453-16	Sequence 16, Appl
30	117	4.9	2780	1 US-08-441-139-1	Sequence 1, Appl1
31	116.5	4.8	4724	4 US-09-066-046-3	Sequence 3, Appl1
32	115.5	4.8	3106	2 US-08-591-079-5	Sequence 5, Appl1
33	115.5	4.8	3695	4 US-09-453-702B-18	Sequence 18, Appl
34	115.5	4.8	6744	1 US-08-119-125A-2	Sequence 2, Appl1
35	114	4.7	1436	4 US-08-961-527-365	Sequence 365, App
36	114	4.7	7766	4 US-09-125-619-3	Sequence 3, Appl1
37	113.5	4.7	4403765	4 US-09-103-840A-2	Sequence 2, Appl1
38	113	4.7	1732	4 US-09-449-335-5	Sequence 5, Appl1
39	112.5	4.7	6270	1 US-08-418-893D-25	Sequence 25, Appl
40	112.5	4.7	6790	1 US-08-418-893D-22	Sequence 22, Appl
41	111.5	4.6	1702	4 US-08-961-527-367	Sequence 367, App
42	111.5	4.6	13842	4 US-09-105-537-30	Sequence 30, Appl
43	111.5	4.6	22671	4 US-08-976-259-14	Sequence 14, Appl
44	111.5	4.6	36778	4 US-09-105-537-5	Sequence 5, Appl1
45	111.5	4.6	38506	3 US-09-320-878-19	Sequence 19, Appl

ALIGNMENTS

RESULT 1
US-08-968-685A-9
Sequence 9, Application US/08968685A
Patent No. 6214981
GENERAL INFORMATION:
APPLICANT: TUCKER, KENNETH
APPLICANT: PLOSLA, LAURA
TITLE OF INVENTION: MORAXELLA CATARRHALIS OUTER MEMBRANE
TITLE OF INVENTION: PROTEIN-106 POLYPEPTIDE, GENE SEQUENCE AND USES THEREOF
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESSES:
ADDRESSEE: PENNIE & EDMONDS LLP
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/968,685A
FILING DATE: No. 6214981ember 12, 1997
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Baldwin, Geraldine F.
REGISTRATION NUMBER: 31,232
REFERENCE/DOCKET NUMBER: 7969-060
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:


```
QY 15 MetGluSerPheArgThrAlaSerAspAlaSerLeuAlaSerSerValArgSerVal 34
      ::||| :::: :::: :::: :::: |||
Db 862 ATGGAATTATATCAGCGACAGCACAAGTAATGTTCAAAACCTAAATACACGGTATGCC 921

QY 35 SerThrThrSerCysArgAspLeuGlnAlaIleThrAspTyrLeuLysHisHisValPhe 54
      |||::: :::: |||||::: |||||
Db 922 AGTTCTATGGGCTCTAGTAATGACAAAGCTGTTCAAGATTATTGAAT----- 969

QY 55 AlaAlaHisArgPheSerVal---IleGlySerProAspGluArgAspAlaAlaLeuAla 73
      ||| :::: ::||| ||||| :::: |||
Db 970 GCAGTAATAATAGTCTTAATGTAGACACTGGAAGCAATAATCAGATCACAATTTCAAAGC 1029

QY 74 HisAsnGluGlnIleAspAlaLeu-----ValGluThrArgAlaAsnArgLeuTyr 90
      ||| ||| ||| ::::| |||||:::
Db 1030 TTGAATATGCAATTTAGAGAATTAGCCTCCACACGCTCAACACGCTGTAATCAAGCCTCT 1089

QY 91 SerGluGlyGluThrProAlaThrIleAlaGluThrPhe----- 103
      ||| ||| ||| ::::| |||||
Db 1090 TCTTTTGA-----GCAGAACTAACCCAAACCTTCAAAAGCATGTCACCTATTTA 1140

QY 104 ----- 106
      ::::|
Db 1141 ATCTCCGGTCTCTTATTCTACGAGCTATCTCTGGACTTAAAGAAATGGTATCCACAGCA 1200

QY 107 GluLysPheAspArgLeuAlaThrThrAlaSerSerAlaPheGluAsnThrProPheAla 126
      ::::| ||| ||| |||
Db 1201 ATAGAAATTTGATCTCATGACGACAAATATTCGCCGTGTATGAATGAGCCGATTTATAA 1260

QY 127 AlaAlaSerValLeuGlnTyrMetGlnProAlaIleAsnLysGlyAspTrpLeuAlaThr 146
      ::::| ::::| ||| ::::| |||||
Db 1261 TATTAATGAACCTTCTC-----CAAGAATCTATTGACTTAGGTGATACACTTTCAAT 1311

QY 147 ProLeuLysProLeuThrProLeuIleSerGlyAlaLeuSerGlyAlaMetAspGlnVal 166
      ::::| ::::| |||
Db 1312 AAAATCACAGATATTCTTCAAAATGACAGCGCATTTTGGAGATGGGTTTCGATGAAGT 1371

QY 167 GlyThrLysMetMetAspArgAlaArgGlyAspLeuHisTyrLeuSer---ThrSerPro 185
      ::::| ::::| ||| ::::| |||||
Db 1372 GAGCTCTCCACGTTAACGAAACTGCCCAAGTTCCTTCAAAATGTCCTGATTAACTCCC 1431

QY 186 AspLysLeuHisAspAlaMetAlaValSerValLysArgHisSerProAlaLeuGlyArg 205
      ||| ::::| ::::| |||
Db 1432 GATGATACAGTTAACACTCTAACGGCAGCAATGCTCAACTTTTATATTGACAGCAAAATGAT 1491

QY 206 GlnVal-----ValAspMetGlyIleAlaValGlnThrPhe 217
      ::::| |||||
Db 1492 TCAATATCAATTGCAGATAAATTAAATGAGGTGATATACTATGCTGTACAACTCTA 1551

QY 218 SerAlaLeuAsnValValArgThrValLeuAlaProAlaLeuAlaSerArgProSerVal 237
      ||| ::::| ||| ::::| |||||
Db 1552 GATCTGGCCAATTCTATCCGTA-----GCTGGTTCACCTGCTTACATTCCGGGTA 1605

QY 238 GlnGlyAlaValAspPheGlyValSerThrAla-----Gly 249
      ::::| ||| ::::| |||||
Db 1606 GAGCTAAATGATCTTATTGGTTATACAACTGCAATTGCTAGTACAAACACGTAATCAGGG 1665

QY 250 GlyLeuValAlaAsn-----AlaGlyPheGlyAspArgMetLeuSer 263
      ::::| ||| ||| |||:::
Db 1666 AATATCGTCGGGAACTCTTAAGACAAATTTTCGCGCGGATGGGAATATCAAAAGCTCA 1725

QY 264 ValGlnSerArgAspGlnLeuArgGlyGlyAlaPheValLeuGlyMetLysAspLysGlu 283
      ::::| ::::| ||| ||| ::::|
Db 1726 ATTAAAGCGTTAGAACAGATGGTATCTCAGTTAAACACAGCTGGTGAAGCTAAATCA 1785

QY 284 ProLysAlaAlaLeuSerGlu-----GluThrAspTrpLeu---AspAlaTyrLys 299
      ::::| ||||| ::::| ||| ||| |||||
Db 1786 GCAAGTGATTTAATTAGTGAAGTTCGTGTAAGTGGGATACGCTTCTGATGCTCAGAAA 1845

QY 300 AlaIleLysSerAlaSerTyrSerGly----- 308
      ||| ::::|
Db 1846 CAAATACTCTCAATTGGAGTAGCTGGTATTATCATATATCCCGTTTAAATGCAATGATG 1905
```

```
QY 309 -----AlaAlaLeuAsnAlaGlyLys----- 315
      ||| ||||| |||
Db 1906 AACAACTTCTCTATTTGCTCAGAAATGCCGCTAAAAACTGCCGCTAACTCAACAGAACTGCT 1965

QY 315 ----- 315
      ::::|
Db 1966 TGGAGTGACAGCAAAAGATATGCAGATAGTCTACAAAGCTAGGGTAATAAGCTTCAAAAT 2025

QY 315 ----- 315
      ::::|
Db 2026 AACTTCAGTGAATTTGCTATTGACAGCTTCTGATGCTTTTATTAGCGAGGATTAATTGAA 2085

QY 316 -----ArgMetAlaGlyLeuProLeuAspValAlaThrAspGlyLeuLysAlaVal--- 332
      ::::| ||||| ||::: ::::|
Db 2086 TTTACTCAAGCCGACAGTTCTTTGGCTTAACGCTTCTACAGAGAGTAATCAATCAGTTGGG 2145

QY 333 -----ArgSerLeuValSerAlaThrSer-----LeuThrLysAsnGly 345
      |||::: ||||| |||
Db 2146 TTCCFACCTCCCTTTTAGCTGACGTAGCACTGCAACCCCTTTGCTCAGTAAGAATACC 2205

QY 346 LeuAlaLeuAlaGly-----GlyTyrAlaGlyValSerLysLeuGlnLysMet 361
      ||||| ||| ::::|
Db 2206 CGCACATTAGCCACAGCAGCCTAATTTGGGCACACGTCGCAATGGGGCAA---GAAACTTTA 2262

QY 362 AlaThrLysAsnIleThrAspSerAlaThrLysAlaAlaValSerGln----- 377
      ||||| ::::| ||::: ||||| |||||
Db 2263 GCGACTGCTGGGCTAGAAAGCTGATGACTCGTGACAGCAGTCCGCTCAAGAGTTCTAAAA 2322

QY 378 -----LeuSerAsnLeuValGlySerValGlyValPheAlaGly 390
      ::::| ||||| |||
Db 2323 ACTGCTCTCGAGGGTTCCTTGTTCACTTTAGTTGGCGGTGCAATTTGCTGCTTGGGA 2382

QY 391 TrpThrThrAlaGlyLeuAlaThrAspProAla---ValLysLysAla----- 405
      ||| ||| ::::| ||| ||||| |||
Db 2383 TGGGCGCTAGAATCATTAATTTCTTCTTTGACAGAGCTAAAAAAGCTAAAGATGATTTT 2442

QY 406 -----GluSer 407
      ::::|
Db 2443 GAGCAGAGCCAGCAAAACCAATGTGAGACAAATTACGACCAATAAAGACTCCACTGATAAA 2502

QY 408 PheIleGln-----AspLysValLysSerThrAlaSerSerThrThrSer 422
      ||||| ||||| ||||| |||
Db 2503 CTAATACAGCAATATTAAGAGCTTCAAAAAGTTAAAGAGTCAAGATCTTTAACTTCAGAT 2562

QY 423 TyrValAlaAspGln-----ThrValLysLeuAlaLysThr-----Val 435
      ::::| ||| ::::| |||||
Db 2563 GAAGAGCAAGAAATACCTTCAAGTCACTACGCAATTAGCACAAACTTTCCTGCATTAGTT 2622

QY 436 Lys-----AspMetSerGlyGluAlaIleSerSerThrGlyAlaSerLeuArgSerThr 453
      ||| ||| ||| ||||| |||
Db 2623 AAAGGCTATGATTTCTCAAGGAAATGCCAATTTCTTAAGACAAATTAAGAGCTTGAAAAAGCG 2682

QY 454 ValAsnAsnLeuArg 458
      ::::| ||| ::::|
Db 2683 ATTGAGAAATACTAAA 2697

RESULT 3
US-08-194-290-6
; Sequence 6, Application US/08194290
; Patent No. 5500353
; GENERAL INFORMATION:
; APPLICANT: Smit, John
; APPLICANT: Bingle, Wade H
; TITLE OF INVENTION: Bacterial surface protein expression
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shlesinger, Arkwright & Garvey
; STREET: 3000 South Eads Street
; CITY: Arlington
; STATE: Virginia
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
```

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/194,290
; FILING DATE: 09-FEB-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Garvey, George A
; REGISTRATION NUMBER: 17737
; REFERENCE/DOCKET NUMBER: 5946-1
; TELEPHONE: 703-684-5600
; TELEFAX: 703-836-5288
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3300 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Caulobacter crescentus
; STRAIN: CB 15
; US-08-194-290-6

Alignment Scores:
Pred. No.: 0.000602 Length: 3300
Score: 132.00 Matches: 101
Percent Similarity: 36.51% Conservative: 75
Best Local Similarity: 20.95% Mismatches: 166
Query Match: 5.48% Indels: 140
DB: 1 Gaps: 23

US-09-825-414-66 (1-487) x US-08-194-290-6 (1-3300)

QY 22 SerAspAlaSerLeuAlaSerSerValArgSerValSerThrThrSerCysArgAsp 41
Db 224 TCGGACGCGCGTGGCTGACCAACACCTGAGCTGTCACACGACGACGCGCTTGGCC 283
QY 42 LeuGlnAlaIleThrAspTyrLeuLysHisHisValPheAlaAlaHisArgPheSerVal 61
Db 284 ATCCAGACC-----TACCAGTCTCTCACCGCGCTTGGCCCGCGGCC 325
QY 62 IleGlySerProAspGluArgAspAlaAlaLeuAlaHisAsnGluGlnIleAspAlaLeu 81
Db 326 GCTGGT-----CTGGACTTCCCTG 343
QY 82 ValGluThrArgAlaAsn-----ArgLeuTyrSerGluGlyGluThr 95
Db 344 GTCGACTCGACCAACCAACCAACGACCTGAACGACGCGTACTACTCGAAG----- 394
QY 96 ProAlaThrIleAlaGluThrPheAlaLysAlaGluLysPhe-----AspArg 111
Db 395 -----TTCGCTCAGGAACCGCTTCATCACTTCTCGATCAAC 433
QY 112 LeuAlaThrThrAlaSerSerAlaPheGluAsnThrProPheAlaAla----- 128
Db 434 CTGGCCACGCGCGCGCGCGCGC-----GCGACGGCTTTCGCCGCGCTACACGGGC 487
QY 129 ---SerValLeuGlnTyrMetGlnProAlaIleAsnLys-----GlyAspTyrPheAla 145
Db 488 GTTTCGTACGCCACAGAGGTGCGCCACCGCTATGACACATCATCGGCAACGCCGTCGCG 547
QY 146 ThrProLeuLysProLeuThrProLeuIleSerGlyAlaLeuSerGlyAlaMetAspGln 165
Db 548 ACC-----GCCGCTGGCGTTCACGCTC 568
QY 166 ValGlyThrLysMetMetAspArgAlaArgGlyAspLeuHisTyrLeuSerThrSerPro 185
Db 569 GCGGCGCGCGTGGCTTTCCTGAACGCGCCAGGCCAACATCGACTACCTGACC----- 619

QY 186 AspLysLeuHisAspAlaMetAlaValSerValLysArgHisSerProAlaLeuGlyArg 205
Db 620 -----GCCCTTCGTGCGGCCCAACACGCGCTTACGGCGCT 655
QY 206 GlnValValAspMetGlyIleAlaValGlnThrPheSerAlaLeuAsnValValArgThr 225
Db 656 GCCGACATCGATCTGGCCCTCAAGCGCCCGCTGATCGGCACCATCTCTGAACGCCGCCACG 715
QY 226 ValLeu-----AlaProAlaLeuAlaSerArgProSerValGlnGlyAlaValAsp 242
Db 716 GTGTGGGCATCGGTGTACCGCAGCCGCCACGGCCGATGATCAACGACCTGTCCGAG 775
QY 248 PheGlyValSerThr-----AlaGlyGlyLeu-----ValAlaAsnAla 255
Db 776 GCGCGCCCTGTGACCCGACACGCGGCTGGCGTGAACCTGTACCGCTATCCGTCTCG 835
QY 256 GlyPheGlyAspArgMetLeuSerValGlnSer---ArgAspGlnLeuArgGly----- 272
Db 836 GCGGTGTGGGTTCGACCTCTCGCTGACCACCGCAGCACCCTGACGGGCGACCGCC 895
QY 273 -----GlyAlaPheValLeuGlyMetLysAsp 281
Db 896 AACAAACGACAGCTTCGTTGCGGGGTGAAGTCCGCCGCTCGACCCCTGACCGTTGGCGAC 955
QY 282 LysGluProLysAlaAlaLeuSerGluGluThrAspTyrProLeuAspAlaTyrLysAlaIle 301
Db 956 ACCCTGAGCGCGGTGTCGACCGACCGCTCCTGACTGGGTGCAAGCT----- 1003
QY 302 LysSerAlaSerTyrSerGlyAlaAlaLeuAsnAlaGlyLysArgMetAlaGlyLeuPro 321
Db 1004 -----GCTGCGGTACGGCT-----CTGCCG 1024
QY 322 LeuAspValAlaThrAspGlyLeuLysAlaValArgSerLeuValSerAlaThrSerLeu 341
Db 1025 ACCGGCGTGACGATCTCGGCATCGAAACGATG---AACGTGACGTGGGGCGCTGCATC 1081
QY 342 ThrLysAsnGlyLeuAlaLeuAlaGlyGlyTyrAlaGlyValSerLysLeu----- 358
Db 1082 ACCCTGAAC-----ACGTCTTCGGGCGTACGGGCTGTGACCGCCCTGAACACCAAC 1132
QY 359 -----GlnLysMetAlaThrLysAsn 365
Db 1133 ACCAGCGCGCGCTCAACCGTCAACCGCGCGCGCTGCGCCAGAACCTGACCGCCACGACC 1192
QY 366 IleThrAspSerAlaThrLysAlaAlaValSerGlnLeuSerAsnLeu---ValGlySer 384
Db 1193 GCGGCTCAAGCCGGAACAACGTGCGCGCTGACGGGGCGCCCAACGTACCCGTCGCTCG 1252
QY 385 ValGlyValPheAlaGlyTyrThrThrAlaGlyLeuAlaThrAspProAlaValLysLys 404
Db 1253 ACGGGCGTGACCTCGGGCAGCACCGGTGCGC-----GCCAACTCGGCC 1297
QY 405 AlaGluSerPheIleGlnAspLysValLysSerThrAlaSerSerThrThrSerTyrVal 424
Db 1298 GCTTCGGGCACCGCTGTCGGTGAGCGTCCGCACTCGACACGACACCAACGCGGCTATC 1357
QY 425 AlaAspGlnThrValLysLeuAlaLysThrValLysAspMetSerGlyGlnAlaIleSer 444
Db 1358 GCC---GTGACCGGTGTACGGCCGTGACCGTGGCTCAAAAGCGCGCAACCGCGTGAAC 1414
QY 445 SerThr 446
Db 1415 ACCACG 1420

RESULT 4
US-08-614-377A-6
; Sequence 6, Application US/08614377A
; Patent No. 5976864
; GENERAL INFORMATION:
; APPLICANT: Smit, John
; APPLICANT: Bingle, Wade H.
; APPLICANT: No. 5976864ellini, John F.
; TITLE OF INVENTION: EXPRESSION AND SECRETION OF

: TITLE OF INVENTION: HETEROLOGOUS
: TITLE OF INVENTION: POLYPEPTIDES FROM CAULOBACTER
: NUMBER OF SEQUENCES: 12
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Fish & Richardson PC
: STREET: 225 Franklin Street
: CITY: Boston
: STATE: Massachusetts
: COUNTRY: USA
: ZIP: 02110-2804
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/614,377A
: FILING DATE: 12-MAR-1996
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/194,290
: FILING DATE: 09-FEB-1994
: CLASSIFICATION: 435
: APPLICATION NUMBER: US 07/895,367
: FILING DATE: 09-JUNE-1992
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Tsao, Y. Rocky
: REGISTRATION NUMBER: 34053
: REFERENCE/DOCKET NUMBER: 08106/002001
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 617-542-5070
: TELEFAX: 617-542-8906
: INFORMATION FOR SEQ ID NO: 6:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 3300 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: ORIGINAL SOURCE:
: ORGANISM: Caulobacter crescentus
: STRAIN: CB 15
: US-08-614-377A-6
Alignment Scores:
Pred. No.: 0.000602 Length: 3300
Score: 132.00 Matches: 101
Percent Similarity: 36.51% Conservative: 75
Best Local Similarity: 20.95% Mismatches: 166
Query Match: 5.48% Indels: 140
Gaps: 23
US-09-825-414-66 (1-487) x US-08-614-377A-6 (1-3300)
QY 22 SerAspAlaSerLeuAlaSerSerValArgSerValSerThrThrSerCysArgAsp 41
Db ||||| ||||| : : : : : ||||| : : : : :
Db 224 TCGAGCGCGCTGGCTGACCAACACCCCTGAAGCTGTCAACAGCAGCAGCGGTGTGCC 283
QY 42 LeuGlnAlaIleThrAspTyrLeuLysHisHisValPheAlaAlaHisArgPheSerVal 61
Db : : : : : : : : : : : ||| : : : : :
Db 284 ATCCAGACC-----TACCAGTTCTTCACCGGCGTGGCCCGCTCGGCC 325
QY 62 IleGlySerProAspGluArgAspAlaAlaLeuAlaHisAsnGluGlnIleAspAlaLeu 81
Db ||| : : : : : : : : : : : : : : : |||
Db 326 GCTGGT-----CTGGACTTCCTG 343
QY 82 ValGluThrArgAlaAsn-----ArgLeuTyrSerGluGluGluThr 95
Db ||| : : : : : ||| : : : : : ||||| : : :
Db 344 GTCGACTGCACCAACACCAACAGACCTGAACGACGCGTACTACTCGAAG----- 394
QY 96 ProAlaThrIleAlaGluThrPheAlaLysAlaGluLysPhe-----AspArg 111
||| ||| : : : |||

Db 395 -----TTGCGCTCAGAAAAACCGCTTCATCAACTTCGATCAAC 433
QY 112 LeuAlaThrThrAlaSerSerAlaPheGluAsnThrProPheAlaAlaAla----- 128
Db ||||| ||| : : : ||||| |||||
Db 434 CTGGCCACGGCGCGCGCGCGG-----GCGAGCGCTTTCGCGCGCGCTACACGGGC 487
QY 129 ---SerValLeuGlnTyrMetGlnProAlaIleAsnLys-----GlyAspTyrLeuAla 145
Db ||| ||| : : : ||| : : : |||
Db 488 GTTCGTACGCCACAGCGGTGCCACCGCCTATGACAAGATCATCGGCACCGCGCTCGCG 547
QY 146 ThrProLeuLysProLeuThrProLeuIleSerGlyAlaLeuSerGlyAlaMetAspGln 165
Db ||| : : : : : : : : : : : : : : :
Db 548 ACC-----GCCGTGGCGTCGACGTC 568
QY 166 ValGlyThrLysMetMetAspArgAlaArgGlyAspLeuHisTyrLeuSerThrSerPro 185
Db GCGGCGCGCGCTTCTCCTGAGCCGCGCACGACCATCGACTACCTGACC----- 619
QY 186 AspLysLeuHisAspAlaMetAlaValSerValLysArgHisSerProAlaLeuLysArg 205
Db ||||| : : : : : : : : : : :
Db 620 -----GCCCTCGTGGCGCCCAACACAGCCGTTACGGCGCGCT 655
QY 206 GlnValValAspMetGlyIleAlaValGlnThrPheSerAlaLeuAsnValValArgThr 225
Db : : : : : : : : : : : : : : :
Db 656 GCCGACATCGATCTGGCGCTCAAGCGCGCGCTGATCGGCACCATCTCTGAACCGCGCACG 715
QY 226 ValLeu-----AlaProAlaLeuAlaSerArgProSerValGlnGlyAlaValAsp 242
Db ||| ||| ||| : : : : :
Db 716 GTGTGGGCATCGGTGTGTACCGGACCGCGCACGCGCGGATGATCAACGACCTGTGGAC 775
QY 243 PheGlyValSerThr-----AlaGlyGlyLeu-----ValAlaAsnAla 255
Db : : : ||||| ||| ||| : : : : :
Db 776 GCGCGCCTGTGACCGACACACGCGCTGCGGTGAACCTGTTCACCGCCTATCCGTGTCG 835
QY 256 GlyPheGlyAspArgMetLeuSerValGlnSer---ArgAspGlnLeuArgGly----- 272
Db ||| ||||| : : : : : ||| ||| |||
Db 836 GCGGTGTGGGTTCGACCTCTCGCTGACCACCGGACACACCTGTACGGGACCGCC 895
QY 273 -----GlyAlaPheValLeuGlyMetLysAsp 281
Db AACACGACACGTTCTGCGGGTGAAGTCGCCGCGCTGCGACCTGACCGGTGGCGAC 955
QY 282 LysGluProLysAlaAlaLeuSerGluGluThrAspTyrPheLysAlaTyrLysAlaIle 301
Db ACCCTGAGCGCGGTGTGGACCGGACGTCCTGAACCTGGGTGCAAGCT----- 1003
QY 302 LysSerAlaSerTyrSerGlyAlaAlaLeuAsnAlaGlyLysArgMetAlaGlyLeuPro 321
Db ||||| : : : : : |||||
Db 1004 -----GCTGCGGTTCAGGCT-----CTGCCG 1024
QY 322 LeuAspValAlaThrAspGlyLeuLysAlaValArgSerLeuValSerAlaThrSerLeu 341
Db ||| ||| : : : : : : : : : : :
Db 1025 ACCGGCGTGACGATCTCGGGCATCGAAGCATG---AACGTGACGTGCGGGCGTGGCATC 1081
QY 342 ThrLysAsnGlyLeuAlaLeuAlaGlyGlyTyrAlaGlyValSerLysLeu----- 358
Db ||| ||| : : : ||| : : : : : |||
Db 1082 ACCCTGAAC-----ACGTCTTCGGGCGGTGACGGGTCTGACCGCCTGAACACCAAC 1132
QY 359 -----GlnLysMetAlaThrLysAsn 365
Db ||| : : : : : : : : : : :
Db 1133 ACCAGCGCGCGGCTCAACCGCTACCGCGCGCGCTGGCCAGAACCTGACCGCCACGAC 1192
QY 366 IleThrAspSerAlaThrLysAlaAlaValSerGlnLeuSerAsnLeu---ValGlySer 384
Db : : : : : ||||| : : : : : ||| |||
Db 1193 GCGGCTCAAGCGCGGACAAACGTGCGGTGACGGCGCGCCAAAGTACCGTGCCTCG 1252
QY 385 ValGlyValPheAlaGlyTyrThrThrAlaGlyLeuAlaThrAspProAlaValLysLys 404
Db ||||| : : : ||| ||||| |||
Db 1253 ACGGCGGTGACCTGCGGACGACACAGGTGCG-----GCCAACTCGGCC 1297
QY 405 AlaGluSerPheIleGlnAspLysValLysSerThrAlaSerSerThrThrSerTyrVal 424
Db ||| : : : ||| : : : : : : : : : : :
Db 1298 GCTTCGGGACACCGGTGTGAGCGTCCGGAACTCGACGACGACACGACGACGCGGCTATC 1357

QY 425 AlaAspGlnThrValLysLeuAlaLysThrValLysAspMetSerGlyGluAlaIleSer 444
Db 1358 GCC--GTGACCGGTGTGTACGGCCGTGACCGGTGGCTCAACAGCGCGGCAACGCCGTGAAC 1414

QY 445 SerThr 446
Db 1415 ACCACG 1420

RESULT 5

US-09-142-648B-6
; Sequence 6, Application US/09142648B
; Patent No. 6210948
; GENERAL INFORMATION:
; APPLICANT: Smit, John
; APPLICANT: Bingle, Wade H.
; APPLICANT: No. 6210948bellnl, John F.
; TITLE OF INVENTION: EXPRESSION AND SECRETION OF HETEROLOGOUS POLYPEPTIDES FROM CAULO
; FILE REFERENCE: 08106/002002
; CURRENT APPLICATION NUMBER: US/09/142,648B
; PCT/CA97/00167
; PRIOR FILING DATE: 1997-03-10
; PRIOR APPLICATION NUMBER: US 07/614,377
; PRIOR FILING DATE: 1996-03-12
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 3300
; TYPE: DNA
; ORGANISM: Caulobacter crescentus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (101)...(3178)
US-09-142-648B-6

Alignment Scores:

Pred. No.: 0.000602 Length: 3300
Score: 132.00 Matches: 101
Percent Similarity: 36.51% Conservative: 75
Best Local Similarity: 20.95% Mismatches: 166
Query Match: 5.48% Indels: 140
DB: 4 Gaps: 23

US-09-825-414-66 (1-487) x US-09-142-648B-6 (1-3300)

QY 22 SerAspAlaSerLeuAlaSerSerValArgSerValSerThrThrSerCysArgasp 41
Db 224 TCGGACCGCGCTGGCTGACCAACCTGAAGCTGTCAACAGCAGCAGCGGTGTGCC 283

QY 42 LeuGlnAlaIleThrAspTyrLeuLysHisHisValPheAlaAlaHisArgPheSerVal 61
Db 284 ATCCAGACC-----TACCAGTCTTCACCGGGTGGCCCGTCCGCC 325

QY 62 IleGlySerProAspGluArgAspAlaAlaLeuAlaHisAsnGluGlnIleAspAlaLeu 81
Db 326 GCTGGT-----CTGGACTTCCTG 343

QY 82 ValGluThrArgAlaAsn-----ArgLeuTyrSerGluGluGlyGluThr 95
Db 344 GTCGACTGCACCAACACCAACGACCTGAACGACGCGTACTACTCGAAG----- 394

QY 96 ProAlaThrIleAlaGluThrPheAlaLysAlaGluLysPhe-----AspArg 111
Db 395 -----TTCGCTCAGGAAACCGCTTCATCAACTTCTCGATCAAC 433

QY 112 LeuAlaThrThrAlaSerSerAlaPheGluAsnThrProPheAlaAlaIle----- 128
Db 434 CTGGCCACGGGGCGCGCGCGCC-----GCGACGGCTTTCGCCGCCCTTACACGGGC 487

QY 129 ---SerValIleGluInTyrMetGlnProAlaIleAsnLys-----GlyAspTrpLeuAla 145
Db 488 GTTTCGTACGCCACAGACGGTTCGCCACCGCTATGACAAAGATCATCGCAACGCCGTCCGC 547

QY 146 ThrProLeuLysProLeuThrProLeuIleSerGlyAlaIleSerGlyAlaMetAspGln 165
Db 548 ACC-----GCCGCTGGCGTGCAGCTC 568

QY 166 ValGlyThrLysMetMetAspArgAlaArgGlyAspLeuHisTyrLeuSerThrSerPro 185
Db 569 GCGGCGCGCGTGGCTTCTCGAGCCGCCAAGGCCAATCGACTACCTGACC----- 619

QY 186 AspLysLeuHisAspAlaMetAlaValSerValLysArgHisSerProAlaLeuGlyArg 205
Db 620 -----GCCTTCGTGCGCGCCACACGCCGTTTCACGGCGCGCT 655

QY 206 GlnValValAspMetGlyIleAlaValGlnThrPheSerAlaLeuAsnValValArgThr 225
Db 656 GCCGACATCGATCTGGCCGTCAAGCGCCCTGATCGGCACCATCTGAACGCCGCCACG 715

QY 226 ValLeu-----AlaProAlaLeuAlaSerArgProSerValGlnGlyAlaValAsp 242
Db 716 GTGTGGGCATCGGTGTGTACGCGCAGCGCCACGCGCGCGATGATCAACGACCTGTCCGAC 775

QY 243 PheGlyValSerThr-----AlaGlyGlyLeu-----ValAlaAsnAla 255
Db 776 GCGCGCTGTGCACCGACACACGCGCGCTGGCTGAACCTGTTCAACGCCCTATCCGTCTGC 835

QY 256 GlyPheGlyAspArgMetLeuSerValGlnSer---ArgAspGlnLeuArgGly----- 272
Db 836 GCGGTGTGGGTTCGACCCCTCTCGCTGACCAACCGGACCGACCAACCCCTGACGGCGCAC 895

QY 273 -----GlyAlaPheValLeuGlyMetLysAsp 281
Db 896 AACACGACACGTTCTGTCGGGTGAAGTCGCCGCGCTGCCACCTGACCGTTGGCGAC 955

QY 282 LysGluProLysAlaAlaLeuSerGluGluThrAspTrpLeuAspAlaTyrLysAlaIle 301
Db 956 ACCCTGACGGCGGTGTGGCAGCAGCTCTGAACCTGGGTGCAAGCT----- 1003

QY 302 LysSerAlaSerTyrSerGlyAlaAlaLeuAsnAlaGlyLysArgMetAlaGlyLeuPro 321
Db 1004 -----GCTGCGGTTCAGGCT-----CTGCCG 1024

QY 322 LeuAspValAlaThrAspGlyLeuLysAlaValArgSerLeuValSerAlaThrSerLeu 341
Db 1025 ACCGGCGTACGATCTCGGGCATCGAAGCATG---AACGTGACGTGGGCGCTGCATC 1081

QY 342 ThrLysAsnGlyLeuAlaLeuAlaGlyTyrAlaGlyValSerLysLeu----- 358
Db 1082 ACCCTGAAC-----ACGTCTTCGGCGGTGACGGGTCTGACCGCCCTGAACACCAAC 1132

QY 359 -----GlnLysMetAlaThrLysAsn 365
Db 1133 ACCAGCGCGCGGCTCAACCGGTACCGCGCGGCTGGGCCAGAACCTGACCGCCACGACC 1192

QY 366 IleThrAspSerAlaThrLysAlaAlaValSerGlnLeuSerAsnLeu---ValGlySer 384
Db 1193 GCGGCTCAAGCGCGAACAACGTCCGCTGACGGGCGCGCCCAACGTACCGTCCCTCG 1252

QY 385 ValGlyValPheAlaGlyTyrThrThrAlaGlyLeuAlaThrAspProAlaValLysLys 404
Db 1253 ACGGCGGTGACCTCGGGCAGCAGCAGGTCCGC-----GCCAATCGGCC 1297

QY 405 AlaGluSerPheIleGlnAspLysValLysSerThrAlaSerSerThrThrSerTyrVal 424
Db 1298 GCTTCGGGACCCGTGTGCGGTGAGCGTCCGAACCTGAGCAGCACCACGCGGCGCTATC 1357

QY 425 AlaAspGlnThrValLysLeuAlaLysThrValLysAspMetSerGlyGluAlaIleSer 444
Db 1358 GCC--GTGACCGGTGTGTACGGCGGTGACCGGTGCTCAACAGCGCGCAACGCCGTGAAC 1414

QY 445 SerThr 446
Db 1415 ACCACG 1420


```
RESULT 6
US-08-591-079-9
; Sequence 9, Application US/08591079
; Patent No. 5972899
; GENERAL INFORMATION:
; APPLICANT: Zychlinsky, Arturo
; APPLICANT: Chen, Yajing
; TITLE OF INVENTION: Apoptosis Induced by Shigella Ipab
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 2000 Pennsylvania Avenue, NW
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20006-1812
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/591,079
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Livnat, Shmuel
; REGISTRATION NUMBER: 33,949
; REFERENCE/DOCKET NUMBER: 15661-20017.00
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 887-1500
; TELEFAX: (202) 887-0764
; TELEX: 90-4030 MRSNFOERSWSH
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5393 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Salmoneila typhi
; STRAIN: Ty2
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 543..2324
; OTHER INFORMATION: /gene= "sipB"
US-08-591-079-9

Alignment Scores:
Pred. No.: 0.00152 Length: 5393
Score: 131.50 Matches: 112
Percent Similarity: 36.248 Conservative: 79
Best Local Similarity: 21.258 Mismatches: 167
Query Match: 5.468 Indels: 169
DB: 2 Gaps: 26

US-09-825-414-66 (1-487) x US-08-591-079-9 (1-5393)
QY 3 IleasnglnSerAlaGlnInProProGlyValAlaMetGluSerPheArgThrAlaSer 22
Db 2373 ATAAATCCCGCCGCTATTAAATATCATCTCTGTGAGATAAGTTTCACACAGACTTCG 2432
QY 23 AspaLaserLeuAlaSerSerSerValArgSerValSerThrThrSerCysArg----- 40
Db 2433 CAATCCGTTAGCGCTAAAGATATTTCTGAATAGTATTGCTATTAGCAGCAGTAAAGTCAGT 2492
QY 41 Aspleu----- 42
Db 2493 GACCTGGGGTGTAGTCTACACTGAGCGCGCTGCGCCAGGGGTATTAAACGAACCCCC 2552
QY 43 GlnAlaIleThrAspTyrLeuLysHisHisValPheAlaIleHisArgPheSerValIle 62
Db 43 GlnAlaIleThrAspTyrLeuLysHisHisValPheAlaIleHisArgPheSerValIle 62
```

```
Db 2553 GGAACGATCACGCTCTTTTAAAA-----GCCAGTATTCAA 2588
QY 63 GlySerProAspGluArgAspAla---AlaLeuAlaHisAsnGluGlnIleAspAlaLeu 81
Db 2589 AATACCGACATGATCATCAGATTTCGAATGCCCTGGCAATATAT----- 2630
QY 82 ValGluThrArgAlaAsnArgLeuTyrSerGluGluThrProAlaThrIleAlaGlu 101
Db 2631 GTACGACTAAAGCGAATGAGTT-----GTGCAAAACCGAGTTACCGCAG 2675
QY 102 ThrPheAlaLysAlaGluLysPheAspArgLeuAlaThrThrAlaSerSerAlaPheGlu 121
Db 2676 CAGCAGCAGAGAGTCGGAAGTTTTTTCATATTAGCGAATGCTCTCCAGTGCCTT--- 2732
QY 122 AsnThrProPheAlaAlaAsnSerValLeuGlnTyrMetGlnProAlaIleAsnLysGlu 141
Db 2733 ---GCGCTGTGGCTGCCGCAATACGTTAATGCTG-----ACGTTGAACCGAGCT 2780
QY 142 AspThrLeuAlaThrProLeuLysProLeuThrProLeuIleSer----- 156
Db 2781 GAT-----AGCAAACTGCTGTAGTAGTGTGCATTTAGTCAGTTTGATGCAGCTAA 2831
QY 157 -----GlyAlaLeuSerGlyAlaMetAspGln 165
Db 2832 ACGACGGCAAGCTCCATGATGCCGGAAGGAGTGAATGCCGTTGTCCGGTAGTATTTCCAG 2891
QY 166 -----ValGlyThrLysMet-----MetAsp 172
Db 2892 AGCGCGCTTCAGTTGGGATCACTGCGCTGGCGCCAACTGGAATATAAGGGGCTGCAG 2951
QY 173 ArgAlaArgGlyAspLeuHisTyrLeuSerThrSerProAspLysLeu----- 188
Db 2952 AATGAAAGAGCGCGCTTAACATAATATGCCGGAAGATCGATAAATGACCACCTGAAGC 3011
QY 189 HisAspAlaMetAlaValSerValLysArgHisSerProAlaLeuGlyArgGlnVal 208
Db 3012 CACAGTATTAAACGCTGTGAACGGCGCAGAAATAGCGTCAAACTTGCTGTAAGCGCTC 3071
QY 209 AspMetGlyIleAlaValGlnThrPheSerAlaLeuAsnValValArgThrValLeuAla 228
Db 3072 GAT-----TCTCTGAATCGTTAAATATGAAGAAACC----- 3104
QY 229 ProAlaLeuAlaSerArgProSerValGlnGlyAlaValAspPheGlyValSerThrAla 248
Db 3105 -----GGTACCAGATGCCAGCAAAATCTTAATGATGCCAGCGCTTAATCTAAT 3152
QY 249 GlyGlyLeuValAlaAsnAlaGlyPheGlyAspArgMetLeuSerValGlnSerArgAsp 268
Db 3153 GCCGGAACGACGCGCACGGAAGTCTGGTATTAA-----AACAGTAATAA 3200
QY 269 GlnLeuArgGlyGlyAlaPheValLeuGlyMetLysAspLysGluProLysAlaLeu 288
Db 3201 CAAATC-----TCCCTGAACATCAGGCTATTCTG 3230
QY 289 SerGluGluThrAspThrLeuAspAlaTyrLysAlaIleLysSerAlaSerTyrSerCly 308
Db 3231 TCGAAACGCTCTTGAGTCTGTGCAATCCGATATTCGTTGAGCAGACAGATACCATGATATG 3290
QY 309 AlaAlaLeuAsnAlaGlyLys---ArgMetAlaGlyLeuProLeuAspValAlaThrAsp 327
Db 3291 ACCCGAATCGATGCCGCAAGATGCAGATGACGGC----- 3326
QY 328 GlyLeuLysAlaValArgSerLeuValSerAlaThrSerLeuThrLysAsnGlyLeuAla 347
Db 3327 -----GATCTGATTATGAAGACTCAGTCACG----- 3353
QY 348 LeuAlaGlyGlyTyrAlaGlyValSerLysLeuGlnLysMetAlaThrLysAsnIleThr 367
Db 3354 ---GTCGGTGTATTGACGGCGCTCCAGG---CAGTACGCCGCTACTCAG----- 3398
QY 368 AspSerAlaThrLysAlaAlaValSerGlnLeuSerAsnLeuValGlySerValGlyVal 387
Db 3399 ---GAACGTTCCGAGCAGCAAAATTAGCCAGGTGAATAACCGGCTTGCACG----- 3446
```

QY 388 PheAlaGlyTrpThrThraAlaGlyLeuAlaThrAspProAlaValLysLysAlaGluSer 407
Db 3447 -----ACCGCATCGAGAGAGAGCCCGTGAAAGTTCACGTAAATCGACACAGC 3491
QY 408 PheIleGlnAspLysValLysSerThrAlaSerSerThrThrSerTyrValAlaAspGln 427
Db 3492 CTGATTCAGGAATGCTGAAACAATGAGAGC----- 3524
QY 428 ThrValLysLeuAlaLysThrValLysAspMetSerGlyGluAlaIleSerSerThrGly 447
Db 3525 -----ATTAAACCAAGTCGAAAGCATCCGCACCTCGCTATCGCA 3563
QY 448 AlaSerLeuArg-----SerThrValAsnAsnLeuArgHisArgSerAlaProGluAla 465
Db 3564 GGCAATATTCGGCTTAATCTGACAGATCACTATACGCCAT----- 3605
QY 466 AspIleGluGluGlyGlyIle 472
Db 3606 -----CAGGGGGGGATT 3617

RESULT 7

US-09-453-702B-72
; Sequence 72, Application US/09453702B
; Patent No. 6365723
; GENERAL INFORMATION:

APPLICANT: Blattner, Frederick R.

Burland, Valerie
Perna, Nicole T.

Plunkett, Guy
Welch, Rod

TITLE OF INVENTION: No. 6365723el Sequences of E. coli O157
NUMBER OF SEQUENCES: 265

CORRESPONDENCE ADDRESS:

ADDRESSEE: Quarles & Brady
STREET: 1 South Pinckney Street
CITY: Madison

STATE: WI

COUNTRY: US
ZIP: 53701-2113

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch. 1.44mb storage
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Word Perfect 8.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/453,702B

FILING DATE: 03-Dec-1999

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 60/110,955

FILING DATE: 04-DEC-1998

ATTORNEY/AGENT INFORMATION:

NAME: Seay, Nicholas J.

REGISTRATION NUMBER: 27386

REFERENCE/DOCKET NUMBER: 960296.95017

TELECOMMUNICATION INFORMATION:

TELEPHONE: (608) 251-5000

TELEFAX: (608) 251-9166

INFORMATION FOR SEQ ID NO: 72:

SEQUENCE CHARACTERISTICS:

LENGTH: 46819

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

SEQUENCE DESCRIPTION: SEQ ID NO: 72:

US-09-453-702B-72

Alignment Scores:

Pred. No.: 0.0765 Length: 46819
Score: 130.00 Matches: 111
Percent Similarity: 35.89% Conservative: 62

Best Local Similarity: 23.03% Mismatches: 232
Query Match: 5.40% Indels: 78
Db: 4 Gaps: 18
US-09-825-414-66 (1-487) x US-09-453-702B-72 (1-46819)
QY 5 GlnSerAlaGlnGlnProProGlyValAlaMetGluSerPheArgThrAlaSerAspAla 24
Db 32171 CAGTCAGCTGCTTCCAGCGCAGGAACGGCATCAACAAGGCTACTGAAGCATCAAAAAGT 32230
QY 25 SerLeuAlaSerSerValArgSerValSerThrThrSerCysArg----- 40
Db 32231 GCTGCCGCTGCAGAGTCTCAAAAAGCGCGGCTACCAG-TGCCGCTGCCGGAACAC 32289
QY 41 -----AspLeuGlnAlaIleThrAspTyrLeuLysHisValPheAlaAlaHis 57
Db 32290 GTCAGAAACGAAATGCCGGCAGTGTCAACAATCAGCCGCAC-----TCTGCATCCAC 32343
QY 58 Arg-----PheSerVal-IleGlySerProAspGluArgAspAlaAlaAlaHisAs 75
Db 32344 CGCGACCACGAAAGCGTCAGAAAGCTGCTCTCAGCCAGGAGGATGCGTCCGCT--TC 32397
QY 75 nGluGlnIleAspAlaLeuValGluThrArgAlaAsnArgLeuTyrSerGluGlyGluTh 95
Db 32398 AAAAGAGCGCGCAAAATCATCAGAAACGAGCGCAGCCTCGAGCGCGCATGTGCAGCCTC 32457
QY 95 rProAlaThrIleAlaGluThrPheAlaLysAlaGluLysPheAspArgLeu--AlaThr 114
Db 32458 CTCGCAACGGCGGAGGAGCAATTCGCGGAGAGCGGCCAAACGTTGAGACAACGCTAA 32517
QY 114 rThrAlaSerSerAlaPheGluAsnThrProPheAlaAlaAlaSerValLeuGlnTyrMe 134
Db 32518 GTCCCTGGAACGGCAGACAGACAGAGTGCCTCCGACAGCAGCAGC----- 32563
QY 134 tGlnProAlaIleAsnLysGlyAspTyrPheAlaThrProLeuLysProLeuThrProLe 154
Db 32564 -----TCAAAAACAGCGGCTGCATT 32583
QY 154 uIleSerGlyAlaLeuSerGlyAlaMetAspGlnValGlyThrLysMetMetAspArgAl 174
Db 32584 ATCTGCCAGTGCCTGCTCAACAAGTCCGCGGACAGCCTCAGCCAGT----- 32629
QY 174 aArgGlyAspLeuHisTyrLeuSerThrSerProAspLysLeuHisAspAlaMetAlaVa 194
Db 32630 -----GCCACCGCGCGGAAATCGGAAATCGGCAAAAGTCCGCGCATC 32667
QY 194 lSerValLysArgHisSerProAlaLeuGlyArgGlnValValAspMetGlyIleAlaVa 214
Db 32668 GTCTGCTTCAACAGCCACACGAAGGCTGGCGAAGCCACTGAACAGGCCAGCAGCAGC 32727
QY 214 lGlnThrPheSerAlaLeuAsnValValArgThrValLeuAlaProAlaLeuAlaSerAr 234
Db 32728 GAGTTCGCTTCCGACGGAAGACATCCGAACGAAACGCGTGGAAACACAGCGC 32787
QY 234 gProSerValGlnGlyAlaValAspPheGlyValSerThrAlaGlyGlyLeuValAlaAs 254
Db 32788 AGAATCCTCAAAAACGGCTGCCGATGTCAGCCAGTTGCGGCGGCTCATCGGCATCATC 32847
QY 254 nAlaGlyPheGlyAspArgMetLeuSerValGlnSerArgAspGlnLeuArgGlyGlyAl 274
Db 32848 TGGGTCT-----GCTTCAAAAAGATGAGCGCAGCACAGACAAGC 32883
QY 274 apheValLeuGlyMetLysAspLysGluProLysAlaAlaLeuSerGluGluThrAspTr 294
Db 32884 GTCAGCAGCGAAGACAGCAGCGCACGAGCATCCACGAAAGCGCAGAGAGCAGCTGTAG 32943
QY 294 pleuAspAlaTyrLysAlaIleLysSerAlaSerTyrSerGlyAla-----AlaLeuAs 312
Db 32944 TCGGACGCGCAGCAGCTCAGAGCAAAAGTACGGCGGAATCTGCAGCAACGCGCGCTGAGAC 33003
QY 312 nAlaGlyLysArg-----MetAlaGlyLeuProLeuAspValAlaThrAspGl 328
Db 33004 AGCGGCAAAACGGCGCAGAGGATATTGCAATCCGCCGCTTGAGGATGCGAGCAGCAGC 33063

QY 328 yLeuLysAlaValArgSerLeuValSerAlaThrSerLeuThrLysAsnGlyLeuAlaLe 348
Db 33064 GAAAAAGGGGATAGTACAGCTCAGCAGTGGCAGCTAACAGCACTTCCGAGTCACTGGCGGC 33123
QY 348 uAlaGlyGlyTyrAlaGlyValSerLysLeuGlnLysMetAlaThrLysAsnIleThr-- 367
Db 33124 AACGCCA-----AAAGCCGTTAAGCGCCGCTATGAGCTGCTAACGGGAAATACACCGC 33177
QY 368 ---AspSerAlaThr---LysAlaAlaValSerGlnLeuSerAsnLeuValGlySerVa 385
Db 33178 ACAGGATGCACAGCAGCAGCAAGAGGATAGTTCAGCTTAGCAACGCGACCAACAGC-- 33235
QY 385 lGlyValPheAlaGlyTrpThrThrAlaGlyLeuAlaThrAspPro--AlaValLysLy 404
Db 33236 -----ACATCTGAATGCTGGCGGCAACGCCAAAGTCGGTAAAGGC 33276
QY 404 sAlaGlySerPheIleGlnAspLysValLysSerThrAlaSerSerThrThrSerTyrVa 424
Db 33277 AGCCTATGACCTTGCTAACGGGAATAT-----ACTGCTCAGAGCGCTACGACA----- 33325
QY 424 lAlaAspGlnThrValLysLeuAlaLysThrValLysAspMetSerGlyGlnAlaIleSe 444
Db 33326 -GCACAAAAGGAATGTCTCCAGCTCAGTACGTACCAACACAGCGCATCTGAACGCTTGC 33384
QY 444 rSerThrGlyAlaSerLeuArgSerThrValAsnAsnLeuArgHisArgSerAlaProGl 464
Db 33385 CGCGACACCAGAAAGCAGTGAAGCAGCTAATGATTAATGCCAATGTGTCGG--GTACCTTC 33441
QY 464 uAla 465
Db 33442 TGCC 33445

RESULT 8
US-09-103-840A-2/c
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4403765
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; FEATURE:
; OTHER INFORMATION: CDC 1551
; OTHER INFORMATION: "n" bases at various positions throughout the sequence
; OTHER INFORMATION: represent a, t, c or g
US-09-103-840A-2

Alignment Scores:
Pred. No.: 244 Length: 4403765
Score: 127.50 Matches: 125
Percent Similarity: 38.64% Conservative: 86
Best Local Similarity: 22.89% Mismatches: 212
Query Match: 5.30% Indels: 124
DB: 4 Gaps: 27

US-09-825-414-66 (1-487) x US-09-103-840A-2 (1-4403765)
QY 8 GlnGlnProProGlyValAlaMetGlySerPheArgThrAlaSerAspAlaSerLeuAla 27
Db 1261064 GAGCACCACCGGCGTGCAGAGCTGCGAGTCCGAAGACCTTGCACGCCGTGCGGAGCGCGCC 1261005

QY 28 Ser-----SerSerValArgSerValSerThrThrSerCysArg----- 40
Db 1261004 GGGCCCGGGCGGCTCATCGGCTCGTGGATGGGCGGTTCGGCCGTGCTGTGCCGGGTGCA 1260945
QY 40 ----- 40
Db 1260944 TCGGTTGTCACCTCAATGCGATCGGCTGCGCGTCCGGCGGGGTGAACACCCGTCGATC 1260885
QY 41 AspleuGlnAlaIleThrAspTyrLeuLysHisHisValPheAlaAlaHisArgPheSer 60
Db 1260884 GAAGGACAGCAGCTGACCCAGCCTGTACGTGCTCAACCTTTACCGCA-----ACC 1260834
QY 61 ValIleGlySerProAspGluArgAspAlaAlaLeuAlaHisAsnGlnIleAspAla 80
Db 1260833 ATCACCGGCTCCCG-----CGCATCGCGCCCGCGCGGAAGTC----- 1260795
QY 81 leuValGluThrArgAlaAsnArgLeuTyrSerGluGlyGluThrProAlaThrIleAla 100
Db 1260794 -----AAGCGCGCCAGCAAGGCTACTGGCGCGGACGTACAGCCGATCCGAGCTG 1260744
QY 101 GluThrPheAlaLysAlaGluLysPheAspArgLeuAlaThrThrAlaSerSerAlaPhe 120
Db 1260743 GAGGCGGTGCGCGGCAAGCTTACGCGCGACACCTGGTTCGGCCCTGGCGCGCGGTCTG 1260684
QY 121 GluAsnThrProPheAlaAlaAlaSerValLeu---GlnTyrMetGlnProAlaIleAsn 139
Db 1260683 GACTCGGTGCGCGGTGAACACCTTCTCTACTACGACCAATGCTCGATACCGCGGTGCTG 1260624
QY 140 LysGlyAspTrpLeuAlaThrProLeuLysProLeuThrProLeuIleSerGly----- 157
Db 1260623 CTCGGC-----GCCGTGCGCGCCCGA--GTAGACCCGGTTTCGAGCGGCGTGGAC 1260576
QY 158 -----AlaLeuSerGlyAlaMetAspGlnValGly-----ThrLysMet 170
Db 1260575 CGCTATTTCGCGCGGCGGGGCGACCGACCGACAGATGCGCGCTGGAGATGACGGAAGTGG 1260516
QY 171 MetAspArgAlaArgGlyAspLeuHisTyrLeu-----SerThrSerPro 185
Db 1260515 TTCGACACG-----AACTACCACTACCTGGTACCAGATCGGGCGCGTGCACACAG 1260465
QY 186 AspLysLeuHisAspAlaMetAlaValSerValLysArgHisSerProAlaLeuGlyArg 205
Db 1260464 TTCACGCTGCACCCCGGCAAGGTGCTCGCCGCACTCAAGAG-----CGGTAGGGCAA 1260411
QY 206 GlnValValAspMetGlyIleAlaValGlnThrPheSerAlaLeuAsnValValArgThr 225
Db 1260410 GGCAATTCCGCGACGTCGGGTGATCATCGGGCGGATCATCTTCTGCTGAGCAAGGCC 1260351
QY 226 ValLeuAlaProAlaLeuAlaSerArg--ProSerValGlnGlyAlaValAspPheGly 244
Db 1260350 GTCGACGGCGGGGGGG--GCCGATCGAACGCGCTGAAGATTGGTTCCGCTCATTTCCGA 1260292
QY 245 ValSerThrAlaGly-----GlyLeuValAlaAsnAlaGlyPhe 257
Db 1260291 GCTGCTGTGCTGCTGTCGCGACGCGCGCGCGCCACAGTGGGTGACGTTCGACGAGCGCGCT 1260232
QY 258 GlyAspArgMetLeuSerValGlnSerArgAspGlnLeuArgGlyAlaPheValLeu 277
Db 1260231 GGTGACCGACCTCTCCCGGACGCGCGCGCTGCTGAAGCGGT----- 1260187
QY 278 GlyMetLysAspLysGluProLysAlaAlaLeuSerGluGlu-ThrAspTrpLeuAspAl 297
Db 1260186 -----GTACACCGCGGCTGTGCTCGGTGACCAACCG----- 1260156
QY 297 aTyrLysAlaIleLysSerAlaSerTyr-----SerGlyAlaAlaLeuAsnAlaGl 314
Db 1260155 ---CCTGCGATCTATGTCGCCACCTACTTCGGGGGACCCGGGCGGCTTACGGCGG-- 1260102
QY 314 yLysArgMetAlaGlyLeuProLeuAsp--ValAlaThrAspGlyLeuLysAlaValAr 333
Db 1260101 -----CTGGCTCGACCCCGGTGGAAGCATGCGCTCGAC----- 1260066
QY 333 gSerLeuValSerAlaThrSerLeuThrLysAsnGlyLeu--AlaLeuAlaGlyGly-- 351

|||||::: ||::: ||::: ||||| |||||
Db 1260065 ----CTGGTGGCCGGTGCCGACACCTCGGTGGCCGGGGGTACCCGAGCTGGCCGGCAAGAC 1260010
QY 352 ----TyrAlaGlyValSerLysLeuGlnLysMetalThrLysAsnIleThrAspSerAl 370
Db 1260009 GCTGGTGGCCGGGGTTC-----GTCGACGGGGCGCAACGTCGTGGCCGACCGA 1259965
QY 370 aThrLysAlaAlaValSerGlnLeuSerAsnLeuValGlySerValGlyValPheAlaGl 390
Db 1259964 CCTGGAGGGCGGGCTGGGCACGTTGGCGACCTGCTGGGTTCGGCGGCTAACCGTGCCGT 1259905
QY 390 yTrpThrThrAlaGlyLeuAlaThrAspProAlaValLysLysAlaGluSerPheIleGl 410
Db 1259904 CTCGACGTCGTGCTCGACACTGCACGTCGCCGTACTCGCTGAACCGAAACCGACCTGA 1259845
QY 410 nAspLysValLysSerThrAlaSerSerThrThrSerTyrValAlaAspGlnThrVally 430
Db 1259844 TGACGCGTGGGGAGCTGGCTGGCGTTCGGTGCCGAAAGGTGGCG--GAAGTCGTGCT 1259788
QY 430 sLeuAlaLysThrValLysAspMetSerGlyGluAlaIleSerSerThrGlyAlaSerle 450
Db 1259787 TCTCGCGCGTGCCTTCGCGAC--GGACACGACGCGGTCGCCGACGAGATCGCGTCGTC 1259731
QY 450 uArgSerThrVal-----AsnAsnLeuArgHisArgSe 461
Db 1259730 CCGCGCCGCCATCGCGTCCCGCAAGCCGCGCGGTTACACAAATGGCAAAATCCGGGC 1259671
QY 461 rAlaProGluAlaAspIleGluGlyGly-IleSerAlaPheSerArgSerGluThrP 481
Db 1259670 GCGCATCGAGCGATCGTCGCGTCCGGAGCCACCGCGCAATGCCGCCAGCGCGCGC 1259611
QY 481 ropheGlnLeuArg 485
Db 1259610 CAGCCAAGACGCGC 1259597
RESULT 9
US-08-961-527-71
; Sequence 71, Application us/08961527
; Patent No. 6420135
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 391
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,527
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 71:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 32768 base pairs
; TYPE: nucleic acid

; STRANDEDNESS: double
; TOPOLOGY: linear
US-08-961-527-71
Alignment Scores:
Pred. No : 0.101 Length: 32768
Score: 126.50 Matches: 97
Percent Similarity: 36.27% Conservative: 76
Best Local Similarity: 20.34% Mismatches: 249
Query Match: 5.26% Indels: 56
DB: 4 Gaps: 9
US-09-825-414-66 (1-487) x US-08-961-527-71 (1-32768)
QY 14 AlameTGluserPheArgThrAlaSerAspAlaSerLeuAlaSerSerSerValArgSer 33
Db 1508 GCCTCTGAATCAGCATCAACAAGTGCCTTCGAGCAAGCACCAGTCGCTGCA 1567
QY 34 ValSerThr-----ThrSerCysArgAspLeuGlnAlaIleThr 46
Db 1568 GCAAGTACTAGTCATCGGCTTCAGCATCGACAAGTGCCTGTAATCGCATCAACGAGT 1627
QY 47 AspTyrLeuLysHisHisValPheAlaAlaHisArgPheSerValIleGlySerProasp 66
Db 1628 GCTTCGGCTTCAGCATCAACGAGTGCCTGACCTTCAGCAAGCACATCAGCTTCTGAATCT 1687
QY 67 GluArgAspAlaAlaLeuAlaHis--AsnGluGlnIleAspAlaLeuValGluThrArg 85
Db 1688 GCATCAACCAAGTGGTCCGCTTCAGCGTCAACCAAGTGCCTGCGCTTCAGCGTCGACAAGT 1747
QY 86 AlaAsnArgLeuTyrSerGluGlyGluThrProAlaThrIleAlaGluThrPheAlaLys 105
Db 1748 GCTTCGGCTTCAGCATCAACGAGTGCCTGCGCTTCAGCAAGCCGCAAGTACCTCAGCGTCA 1807
QY 106 AlaGluLysPheAspArgLeuAlaThrThrAlaSerSerAlaPheGluAsnThrProPhe 125
Db 1808 GCTTCC-----GCCTCAACCAAGTGCCTGCGCTTCAGCAAGCACACAAGTGC 1852
QY 126 AlaAlaAlaSerValLeuGlnTyrMetGlnProAlaIleAsnLysGlyAspTrpLeuAla 145
Db 1853 TCAGCCTCAGCAAGTATCTCAGCGTCTGAATCGGCATCAACGAGTGCCTGAGTCAGCA 1912
QY 146 ThrProLeuLysProLeuThrProLeuIleSerGlyAlaLeuSerGlyAlaMetaspGln 165
Db 1913 TCA-----ACGAGTACGTCAGCCTCAGCAAGCACATCAGCTTCTGAATCT 1957
QY 166 ValGlyThrLysMetMetaspArgAlaArgGlyAspLeuHisTyrLeuSerThrSer-Pr 185
Db 1958 GCATCAACCAAGTGCCTCAGCCTCAGCA-----TCGACAAAGCGCC 1996
QY 185 oAsp-----LysLeuHisAspAlaMetAl 193
Db 1997 TCAGCTTCAGCAAGTACCAAGTGTTCAGGCTTCAGCGTCGACAAGTGGCGGCTCAACC 2056
QY 193 aValSerValLysArgHisSerProAlaLeuGlyArgGlnValValaspMetGlyIleAl 213
Db 2057 AGTGCATCTGAATCGCATCAACCAAGTGCCTCAGCCTCAGCAAGTACT-AGTGCATCAGC 2115
QY 213 aValGlnThrPheSerAlaLeuAsnValValArgThrValLeuAlaProAlaLeuAlaSe 233
Db 2116 TTCAGCATCAACGAGTGCATCGGCTTCAGCATCAACCAAGTGCCTGCGCTTCAGCGTCAAC 2175
QY 233 rArgProSerValGlnGlyAlaValaValaspheGlyValSerThrAlaGlyGlyLeuValAl 253
Db 2176 CAGTGCCTCAGCTTCAGCAAGTACCAAGTGTTCAGTCTCAGCATCAACAAGTGTTCAGC 2235
QY 253 aAsnAlaGlyPheGlyAspArgMetLeuSerValGlnSerArgAspGlnLeuArgGlyGl 273
Db 2236 CTCAGCATCG-----ACAAGTGCCTCGGCTTC 2262
QY 273 yAlaPheValLeuGlyMetLysAspLysGluProLysAlaAlaLeuSerGluGluThrAs 293
Db 2263 AGCAAGCACATCAGCATCTGAATCAGCGGTCAACCAAGTGTTCGGCTTCAGCAAGTACCAG 2322

QY 293 pTrpLeuAspAlaTyrLysAlaIleLysSerAlaSerTyrSerGlyAlaAlaLeuAsnAl 313
Db 2323 TGCCTCAGCTTCAGCATCAACGAGCGCCCTCGGCTCAGCAAGCACCCTCAGCTTCTGAATC 2382
QY 313 aGlyLysArgMetAlaGlyLeuProLeuAspVal---AlaThrAspGlyLeuLysAlaVala 332
Db 2383 GGCCTCAACGAGCGCTCGGCTCAGCAAGCACCCTCAGCTTCTGAATCGGCTCAACGAG 2442
QY 332 lArgSerLeuValSerAlaThrSerLeuThrLysAsnGlyLeuAlaLeuAlaGlyGlyTy 352
Db 2443 CGCCTCAGCCTCAGCATCAACGAGAGTGTCTCGGCTCAGCAAGCACAAGCGCCCTCGGCTTC 2502
QY 352 rAlaGlyValSerLysLeuGlnLysMetAlaThr-----LysAsnIleThrAs 368
Db 2503 AGCATCAACGAGTACGTACGCTTACAGCGTCAACCAAGTGTCTTCAAGCCTCAGCAATCAACAAG 2562
QY 368 pSerAlaThrLysAlaAlaValSerGlnLeuSerAsnLeuValGlySerValGlyValPh 388
Db 2563 TGCCTCAGCCTCAGCAAGTATCTCAGCGTCTGAATCGGCATCAACGAGTGGCTGAATC 2622
QY 388 eAlaGlyTyrThrAlaGlyLeuAlaThrAspProAlaValLysLysAlaGlySerPh 408
Db 2623 AGCATCAACGAGTACGTACGCTCAGCAAGCACCCTCAGCTTCTGAATCGGCTCAACGAG 2682
QY 408 eIleGlnAspLysValLysSerThrAlaSerSerThrThrSerTyrValAlaAspGlnTh 428
Db 2683 TGCCTCAGCCTCAGCATCAACGAGCGCTCAGCTTCAACGAGTACCAAGTACGCTTCAAGCCTC 2742
QY 428 rValLysLeuAlaLysThrValLysAspMetSerGlyGlyAlaIleSerSerThrGlyAl 448
Db 2743 AGCGTCGACAAGTGGCTGGCGCTCAACGAGTGAATCTGAATCGGCATCAACGAGTGGCTC 2802
QY 448 aSerLeuArgSerThr-ValAsnAsnLeuArgHisArgSerAlaPro 463
Db 2803 AGCCTCAGCAAGTACTAGTGCATCGGCTTCAAGCATCAACGAGTGGCT 2849
RESULT 10
US-09-103-840A-1/c
; Sequence 1, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 4411529
; TYPE: DNA
; ORGANISM: Mycobacterium tuberculosis
; OTHER INFORMATION: H37Rv
US-09-103-840A-1
Alignment Scores:
Pred. No.: 314 Length: 4411529
Score: 126.50 Matches: 125
Percent Similarity: 38.64% Conservative: 86
Best Local Similarity: 22.89% Mismatches: 212
Query Match: 5.26% Indels: 124
DB: 4 Gaps: 27
US-09-825-414-66 (1-487) x US-09-103-840A-1 (1-4411529)
QY 8 GlnGlnProGlyValAlaMetGluSerPheArgThrAlaSerAspAlaSerLeuAla 27
Db 1261536 GAGCACCACCGCGGTGCGAGCCTGCGAGTCCGAGACCTTCCAGCGCGGTGCGAGCGCGC 1261477

QY 28 Ser-----SerSerValArgSerValSerThrThrSerCysArg----- 40
Db 1261476 GCGCCCGGGGCTCATGCGCTCGTGGAAATGGGCGTTTGGCCGTGTCGGCGGTCGA 1261417
QY 40 ----- 40
Db 1261416 TCGGTGTGCACCTCAATCGGATCGGCTGCGCGTCCGCGGCGGTGAACCAACCCGTGATC 1261357
QY 41 AspLeuGlnAlaIleThrAspTyrLeuLysHisHisValPheAlaAlaHisArgPheSer 60
Db 1261356 GAAGGACAGCAGGTGACCCAGCCTGTACGTGCTGAACCTTTACCGCA-----ACC 1261306
QY 61 ValIleGlySerProAspGluArgAspAlaAlaLeuAlaHisAsnGlnIleAspAla 80
Db 1261305 ATCACCAGCTCCCG-----CGCATCGCGCGCGCGCGGAGTGC----- 1261267
QY 81 LeuValGluThrArgAlaAsnArgLeuTyrSerGluGlyGluThrProAlaThrIleAla 100
Db 1261266 -----AAGCGCGCCACCGAAGGCTACTGGCGCGGAGCTACCAAGCCGATCCGAGCTG 1261216
QY 101 GluThrPheAlaLysAlaGlyLysPheAspArgLeuAlaThrThrAlaSerSerAlaPhe 120
Db 1261215 GAGCGCGTCCGCGCCGACGTTACCGCGGACACCTGGTCCGCCCGCGCGGCTCTG 1261156
QY 121 GluAsnThrProPheAlaAlaAlaSerValLeu---GlnTyrMetGlnProAlaIleAsn 139
Db 1261155 GACTCGGTGCGGTGAACACCTTCTCTACTACGACCAATGCTCGATACCGCGGTGCTG 1261096
QY 140 LysGlyAspTrpLeuAlaThrProLeuLysProLeuThrProLeuIleSerGly----- 157
Db 1261095 CTCGGC-----GCCCTGCCCGCCCGCA---GTGAGCCCGGTTTCCGAGCGGCTGGAC 1261048
QY 158 -----AlaLeuSerGlyAlaMetAspGlnValGly-----ThrLysMet 170
Db 1261047 CGCTATTTCGCGCGCGCGGGGACCGACCGACAGATCGCGCGCTGAGATGACGAAGTGG 1260988
QY 171 MetAspArgAlaArgGlyAspLeuHisTyrLeu-----SerThrSerPro 185
Db 1260987 TTCGACAC-----AACTACCACTACTGTTACCCGAGATCGGCGCTGACCAACG 1260937
QY 186 AspLysLeuHisAspAlaMetAlaValSerValLysArgHisSerProAlaLeuGlyArg 205
Db 1260936 TTCACGCTGACCCCGGCAAGGTGCTCGCCGAATCAAGAG-----GCGTTAAGGGCAA 1260883
QY 206 GlnValValAspMetGlyIleAlaValGlnThrPheSerAlaLeuAsnValValArgThr 225
Db 1260882 GGCATTCCCGCACGTCCGTGATCATCGGCGCGATCACTTCTCTGCTGAGCAAGGCC 1260823
QY 226 ValLeuAlaProAlaLeuAlaSerArg---ProSerValGlnGlyAlaValaAspPheGly 244
Db 1260822 GTCGACGGCGCGGGGGC-GCCGATCGAACCGCTCGAAGAGTTGGTTCGGTCTATTTCGGA 1260764
QY 245 ValSerThrAlaGly-----GlyLeuValAlaAsnAlaGlyPhe 257
Db 1260763 GCTGCTGTGCTGCTTGCAGCGGCGGCGCGCCAGTGGGTGCAGTTCGACGAGCGGCGCT 1260704
QY 258 GlyAspArgMetLeuSerValGlnSerArgAspGlnLeuArgGlyGlyAlaPheValLeu 277
Db 1260703 GGTGACCGACCTTCCCGGACGCGCGCCCTGCTGGAAGCGGT----- 1260659
QY 278 GlyMetLysAspLysGluProLysAlaAlaLeuSerGluGlu-ThrAspTrpLeuAspAl 297
Db 1260658 -----GTACACCGCGCTGTGCTCGGTGAGCAACCG----- 1260628
QY 297 aTyrLysAlaIleLysSerAlaSerTyr-----SerGlyAlaAlaLeuAsnAlaGly 314
Db 1260627 -----CCTGCGATCTATGTCCGACACTTTCGGGAGACCCGGCGGCGCTTACCGCGG-- 1260574
QY 314 yLysArgMetAlaGlyLeuProLeuAsp---ValAlaThrAspGlyLeuLysAlaValAr 333
Db 1260573 -----CTGGCTCGACACCCCGGTGGAAGCAATCGGCGTGCAC----- 1260538

```

QY      333  gSerLeuValSerAlaThrSerLeuThrLysAsnGlyLeu--AlaLeuAlaGlyGly-- 351
          |||||:::
Db 1260537  ---CTGGTGGCCGGTGGCCGACACCTCGGTGGCCGGGGTACCAGAGCTGGCCGCAAGAC 1260482
QY      352  ---TyrAlaGlyValSerLysLeuGlnLysMetAlaThrLysAsnIleThrAspSerAl 370
          |||||
Db 1260481  GCTGTGGCCGGGTCTC-----GTCAGCGGGCGCAACGTCGTGGCGCACCGA 1260437
QY      370  aThrLysAlaAlaValSerClnLeuSerAsnLeuValGlySerValGlyValPheAlaG1 390
          ::|||:::
Db 1260436  CCTGGAGCGCGCGTGGGACGCTGGCGACCCCTGCTGGGTTCGGCGGCTACCGTGGCCGT 1260377
QY      390  yTrpThrThrAlaGlyLeuAlaThrAspProAlaValLysLysAlaGluSerPheIleG1 410
          |||:::
Db 1260376  CTCGACGTCGTGCTCGACACTGCACAGCTGCCGTACTCGCTGGAAACCGGAACCGACCTGGA 1260317
QY      410  nAspLysValLysSerThrAlaSerSerThrThrSerTyrValAlaAspGlnThrVally 430
          |||:::
Db 1260316  TGACGCGTGGCGAGCTGGCTGGCGTTCGGTGGTCCGGAAGAGTGGCC--GAAGTGTGCT 1260260
QY      430  sLeuAlaLysThrValLysAspMetSerGlyGluAlaIleSerSerThrGlyAlaSerle 450
          |||||:::
Db 1260259  TCTCGCGCGTGCCTGGCGGAC--GGACACGACGCGGTGCCGACGAGATCGCTGCTC 1260203
QY      450  uArgSerThrVal-----AsnAsnLeuArgHisArgSe 461
          |||:::
Db 1260202  CCGCGCCGCATCGCTCCGCCGAAGCGGACCGCGCGGTTCACACAATGGCAATCCGGGC 1260143
QY      461  rAlaProGluAlaAspIleGluGlyGly-IleSerAlaPheSerArgSerGluThrP 481
          :|||
Db 1260142  GCGCATCGAGCGCATCGCTCCGACGCCACCGCGGCAATGCCGCCAGCGCGCGC 1260083
QY      481  ropheGlnLeuArg 485
          ||:::
Db 1260082  CAGCCAAGACGCGC 1260069

RESULT 11
US-08-682-517-13
; Sequence 13, Application US/08682517
; Patent No. 5874267
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Expression of surface layer proteins
; NUMBER OF SEQUENCES: 25
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/682,517
; FILING DATE:
; CLASSIFICATION:
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3666 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-682-517-13

Alignment Scores:
Pred. No.: 0.00357 Length: 3666
Score: 125.50 Matches: 117
Percent Similarity: 32.99% Conservative: 73
Best Local Similarity: 20.31% Mismatches: 219
Query Match: 5.21% Indels: 167
DB: 2 Gaps: 25

US-09-825-414-66 (1-487) x US-08-682-517-13 (1-3666)
QY      3  IleAsnGlnSerAlaGlnProProGlyValAlaMetGlu----- 16

```

Db	2023	GTAAACCAAGCGGCAAAAAGTAAC	CTGTA	CTTCTTAAGAAAGCACTTATACA	ATC	2082
QY	17	-----	-----	-----	-----	Ser 17
Db	2083	TACAAT	TACTGGTGAATGATATTAAG	TAGATACCAAGTTATCTCACC	AAATCGTAGC	2142
QY	18	Phe	ArgThrAlaSerAspAlaSerLeuAlaSerSerSerValArgSerValSerThr	Thr	37	
Db	2143	TACACAGTA	CTTATGAAGCTACTTATCTTACA	-----	-----	GGA
QY	38	Ser	CysArgAspLeuGlnAlaIleThr	AspTyrLeuLysHisHisValPheAlaAlaHis	57	
Db	2194	CCTGCTTA	GAAATTTAGAGTTACTTCTCAGTGCATGGTAAACAAC	CTGCTGTTAAAGTAAT	2253	
QY	58	ArgPhe	SerValIleGlySerProAsp	GluArgAspAlaAlaLeuAlaHisAsn	Gln	77
Db	2254	GCTACAGG	TATGCTGTATATACAGACGCGTA	AAAGACTATGCATTTACTGCTAAAGAACT	2313	
QY	78	Ile	AspAlaLeuValGluThr	-----	-----	84
Db	2314	ACAGCTAC	ATTACAGCTACAAATGAAGTTC	CAAACTCTTACACTGCTGCTAGCTACTCA	2373	
QY	85	-----	-----	-----	-----	ArgAlaAsnArgLeuTyrSerGluGlyGluThrPro
Db	2374	TTCAAT	TACAGCTGATTCCTGGTTC	CAACACAGCAACTCTATTTGGTTGCTGGTAA	AAACCA	2433
QY	97	Ala	ThrIleAlaGluThrPheAlaLysAlaGluLysPhe	AspArgLeuAlaThrThrAla	116	
Db	2434	GTC	AAATATGCTGGTGATCAGGCCAA	ACATATATAATAC	-----	TTCGAGCTAATGCT
QY	117	Ser	SerAlaPheGluAsnThrProPheAlaAlaAlaSerValLeuGlnTyr	-----	-----	133
Db	2488	AATGA	GTATTTGGTGAAGCGGCATGG	AAGCATTA	-----	TTAACTCAATATGCAACTGAA
QY	134	-----	-----	-----	-----	MetGln 135
Db	2545	GGCCAA	AAAAAGTAACAATCTCATATATG	TAGATGGTGATACAGTTACATTTAAAGTAAT	2604	
QY	136	Pro	AlaIleAsnLysGlyAspTyrPheAlaThrPro	LeuLysProLeuThrProLeu	--- 154	
Db	2605	AGT	GCTGTATATTTCTCAACTGAAGCTATCA	AAACCAAGTTGCTCCAACAACACACAGCAGCT	2664	
QY	155	Ile	SerGlyAlaLeuSer	-----	-----	GlyAlaMetAsp
Db	2665	CCA	ACTACTGGCGCATTTACATTAACAC	CAGCAGCTGGTGGTTAGTTGATTTTACA	ACT 2724	
QY	165	Gln	ValGlyThrLysMetMetAspArgAlaArgGlyAsp	LeuHisTyrLeuSerThrSer	184	
Db	2725	GCA	ACTAACACTTTAGGAATTTCAATTA	GCTGATGCAGATCTTAATGTAAGTGC	AAACACT 2784	
QY	185	Pro	AspLysLeuHisAspAlaMetAlaValSer	ValLys--ArgHisSerProAlaLeu	203	
Db	2785	GTT	-----	-----	-----	GAT
QY	204	Gly	ArgGlnValAlaValAspMetCylleAlaValGln	ThrPheSerAlaLeuAsnValAl	223	
Db	2833	TCT	CTTACATTAAGTGAAGTGGTCTA	TATACAGGTGTATTTGCTACA	-----	2880
QY	224	Arg	ThrValLeuAlaProAlaLeuAlaSerArgPro	SerValGlnGlyAlaValaspPhe	243	
Db	2881	---	ACTGTTCAAGCTGGTACATTA	CTTCT	-----	2907
QY	244	Gly	ValSerThrAlaGlyGlyLeuValAlaAsn	-----	-----	AlaGly 256
Db	2908	---	TTAACTGCTGGTACATTAACAG	TTACTTATGCGAGATGCTAA	AAATGCTGCAGGT 2961	
QY	257	Phe	GlyAspArgMetLeuSerValGlnSerArgAsp	GlnLeuArgGlyGly	-----	273
Db	2962	GTT	GCTGAAAATATATTACCTAGCGT	AACATTAAGAAACTACTGGAGCAAT	TACTTCT 3021	
QY	274	---	AlaPheValLeuGlyMetLysAspLysGluPro	LysAlaAlaLeuSerGluGluThr	292	

Db 3022 GATACATTACACAGGTGTATTA-----CCATCAGCAGCTACAGCAGCTGAATAT 3072
QY 293 AspTrpLeuAspAlaIYrLysAlaIleLysSerAlaSerTyrSerGlyAlaAlaLeu--- 311
Db 3073 ACTCTAAATCAATTCGTCGACAGATTATACATTTCACACAGGTGAAGATTTCACCTTAAAT 3132
QY 312 -----AsnAlaGlyLysArgMetAlaGlyLeuProLeuAspValAlaThrAspGlyLeu 329
Db 3133 ATTGATAATGCTGTCGTCACAGTAATTAACTTAGCA-----GGTAAA 3174
QY 330 LysAlaValArgSerLeuValSerAlaThrSerLeuThrLysAsnGlyLeuAlaLeuAla 349
Db 3175 AAAGGTGCACAGGTGTACGTGATGCTATCAATGCTACATTTCAGAGGTACTGCACTGTT 3234
QY 350 GlyGlyTyrAlaGlyValSerLysLeuGlnLysMetAlaThrLysAsnIleThrAsp--- 368
Db 3235 TCTGGA-----GACAAAGTAGTTATTAAATCAGCTACACAGGT 3273
QY 369 -----SerAlaThrLysAlaAlaValSerGlnLeuSerAsnLeuVal--- 382
Db 3274 GTTGCTTCTGAAGTTGAAGTTACATTCTCTTCGTTAAATCAAGTATTAAATGCAGTAGTT 3333
QY 383 ---GlySerValGlyValPheAlaGlyTrp---ThrThrAlaGlyLeuAlaThrAspPro 400
Db 3334 AACGGTAAAGATCAAGTCGTTGCAGAACAGCAGCTGCACAAAGAATTCACGATTACTACA 3393
QY 401 AlaValLysLysAlaGlySerPheIleGlnAspLysValLysSerThr----- 416
Db 3394 GCCCTTCTGTGGGTGAAGTAGTATTGATGGTGTGAATATACTGCTGTAGCATTT 3453
QY 417 -----AlaSerSerThrThrSerTyr 423
Db 3454 GGAAGCTGCTCCACACAGCAAAATACATTCTGAGTTGAATCTGCTTAATACATTAGCTTCA 3513
QY 424 ValAlaAspGlnThrValLysLeuAlaLysThrValLysAspMetSerGlyGlu----- 441
Db 3514 GTAGCTGACCAAGCTGCAAACTTGTGCTGCTACAAATTGACTTTAAACACTGCAGATTAAG 3573
QY 442 ---AlaIleSerSerThrGlyAla-----SerLeuArgSerThrVal 454
Db 3574 TTTACAGCTTCTGCAACAGCTGCTACTATTACATTAACTTCTACTGTA 3621
RESULT 12
US-08-682-517-14
; Sequence 14, Application US/08682517
; Patent No. 5874267
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Expression of surface layer proteins
; NUMBER OF SEQUENCES: 25
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/682,517
; FILING DATE:
; CLASSIFICATION:
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3666 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..3666
US-08-682-517-14
Alignment Scores: 0.00357 Length: 3666
Pred. No.:

Score: 125.50 Matches: 117
Percent Similarity: 32.99% Conservative: 73
Best Local Similarity: 20.31% Mismatches: 219
Query Match: 5.21% Indels: 167
DB: 2 Gaps: 25
US-09-825-414-66 (1-487) x US-08-682-517-14 (1-3666)
QY 3 IleAsnGlnSerAlaGlnGlnProProGlyValAlaMetGlu----- 16
Db 2023 GTAAACCAAAAGCGCAAAAAGTAAGTAACTGGTACTTCTATTAAAGAAAGCAACTTATACAATC 2082
QY 17 -----Ser 17
Db 2083 TACAATPACTGGTCTAATGATATTAAAGTAGATAACCAAGTTATCTCACCAAATCGTAGC 2142
QY 18 PheArgThrAlaSerAspAlaSerLeuAlaSerSerSerValArgSerValSerThrThr 37
Db 2143 TACACAGTAACCTTATGAAGCTTATCTTACTTACA-----GGAAGCTGTATTACA 2193
QY 38 SerCysArgAspLeuGlnAlaIleThrAspTyrLeuLysHisHisValPheAlaAlaHis 57
Db 2194 CCTGCTAAGAATTTAGAACGTACTTCAAGTGATGGTAAACAACAACCTGCTGTTAAAGTAATT 2253
QY 58 ArgPheSerValIleGlySerProAspGluArgAspAlaAlaLeuAlaHisAsnGln 77
Db 2254 GCTACAGGTATTGCTGTTAATACAGACGGTAAGACATATGCATTACTGCTAAGAACST 2313
QY 78 IleAspAlaLeuValGlyThr----- 84
Db 2314 ACACTACATTCCACAGCTACAAATGAAGTTCCAAACTCTTACACTGGTGTAGCTACTCAA 2373
QY 85 -----ArgAlaAsnArgLeuTyrSerGlyGlyThrPro 96
Db 2374 TTCAATACAGCTGATTCGTGTTCAAAACAGCAACCTGTATTGGTTGCTGTAACCAACA 2433
QY 97 AlaThrIleAlaGlyThrPheAlaLysAlaGlyLysPheAspArgLeuAlaThrThrAla 116
Db 2434 GTGAATATGCTGCTGATGACGCAAAACATATAAATAC-----TTGAGAGCTAAATGGT 2487
QY 117 SerSerAlaPheGluAsnThrProPheAlaAlaAlaSerValLeuGlnTyr----- 133
Db 2488 AATGAAGTATTGTTGTAAGCGGCATGGGAAGCATTA--TTAACTCAATATGCAACTGAA 2544
QY 134 -----MetGln 135
Db 2545 GGCCAAAAGTAACAATCTCATATAATGTAGATGGTGATACAGTTACATTAAAGTAAAT 2604
QY 136 ProAlaIleAsnLysGlyAspTrpLeuAlaThrProLeuLysProLeuThrProLeu--- 154
Db 2605 AGTCTGTTAATTCTTCAACTGAAGCTATCAACACAGTTGCTCCAACAACACAGCAGCT 2664
QY 155 ---IleSerGlyAlaLeuSer-----GlyAlaMetAsp----- 164
Db 2665 CCAACTACTGGCGCATTTAACAACACAGCAGCTGGTGGTTTATGATTGATTAAACAAC 2724
QY 165 GlnValGlyThrLysMetMetAspArgAlaArgGlyAspLeuHisTyrLeuSerThrSer 184
Db 2725 GCAACTAACACTTTAGGAATTTCAATTAGCTGATGCAGATCTTAATGTAAGTCAACAAC 2784
QY 185 ProAspLysLeuHisAspAlaMetAlaValSerValLys---ArgHisSerProAlaLeu 203
Db 2785 GTT-----GATACTGCAACTGTTTCATTTAAAGATAGTGCAAATATATTCATT 2832
QY 204 GlyArgGlnValValAspMetClyIleAlaValGlnThrPheSerAlaLeuAsnValVal 223
Db 2833 TCTCTTACATTAGTTGAACCTGTGCTAATACAGGTGTATTGCTACA----- 2880
QY 224 ArgThrValLeuAlaProAlaLeuAlaSerArgProSerValGlnGlyAlaValAspPhe 243
Db 2881 ---ACTGTTCAAGCTGTGATCATTTATCTTCT----- 2907
QY 244 GlyValSerThrAlaGlyGlyLeuValAlaAsn-----AlaGly 256

```

Db      2908  -----TTAACTGCGGTGATACATTAAACAGTACTTATGACAGATGCTAAAAAATGCTGCAGGT 2961
              ||| ||| ||| ||| |||
QY      257  PheGlyAspArgMetLeuSerValGlnSerArgAspGlnLeuArgGlyGly----- 273
              ::: ::: ::: ::: :::
Db      2962  GTTGCTGAAATATTTACTGCTAGCGTACATTAAGCAAAACTACTGGAGCAATTACTTCT 3021
              ::: ::: ::: :::
QY      274  ---AlaPheValLeuGlyMetLysAspLysGluProLysAlaAlaLeuSerGluIuThr 292
              ||| ||| ||| ||| |||
Db      3022  GATACATTACACAGGTGTATTA-----CCATCAGCAGCTACAGCAGCTGAATAT 3072
              ||| ||| ||| ||| |||
QY      293  AspTrpLeuAspAlaTyrLysAlaIleLysSerAlaSerTyrSerGlyAlaAlaLeu--- 311
              ||| ||| ||| ||| |||
Db      3073  ACTTCTAAATCAATTCTGCTGCAGATTATACATTTCGCAACAGGTGAAGGATTCACTTTAAAT 3132
              ||| ||| ||| ||| |||
QY      312  -----AsnAlaGlyLysArgMetAlaGlyLeuProLeuAspValAlaThrAspGlyLeu 329
              ||| ||| ||| ||| |||
Db      3133  ATTGATTAATGCTGGTCTCAAGTAATTAACCTTAGCA-----GGTAAA 3174
              ||| ||| ||| ||| |||
QY      330  LysAlaValArgSerLeuValSerAlaThrSerLeuThrLysAsnGlyLeuAlaLeuAla 349
              ||| ||| ||| ||| |||
Db      3175  AAAGGTGCACAAAGGTGTAGCTGATGCTATCAATGCTACATTTGCAGGTACTGCACACTGTT 3234
              ||| ||| ||| ||| |||
QY      350  GlyGlyTyrAlaGlyValSerLysLeuGlnLysMetAlaThrLysAsnIleThrAsp--- 368
              ||| ||| ||| ||| |||
Db      3235  TCTGGA-----GACAAAGTAGTTATTAAATCAGCTACACACAGGT 3273
              ||| ||| ||| ||| |||
QY      369  -----SerAlaThrLysAlaAlaValSerGlnLeuSerAsnLeuVal--- 382
              ||| ||| ||| ||| |||
Db      3274  GTTGGTTCTGAAGTTGAAGTTACATTCTCTCTGTTAATCAAGTATTAAATGCAGTAGTT 3333
              ||| ||| ||| ||| |||
QY      383  ---GlySerValGlyValPheAlaGlyTrp---ThrThrAlaGlyLeuAlaThrAspPro 400
              ||| ||| ||| ||| |||
Db      3334  AACGGTAAGATCAAGTCGTTGCAGGACAGCTGCTACAAAAGCATTCACGATTACTACA 3393
              ||| ||| ||| ||| |||
QY      401  AlaValLysLysAlaGlnSerPheIleGlnAspLysValLysSerThr----- 416
              |||::: ||| ::: ||| |||::: |||
Db      3394  GCCCTTCTGTGGTGA AAAAGTAGTTATGATGCTGTAATATACCTGCTGACATTT 3453
              |||::: ||| ::: ||| |||::: |||
QY      417  -----AlaSerSerThrThrSerTyr 423
              |||::: ||| ::: ||| |||::: |||
Db      3454  GGAAGTGTCTCCAACAGCAATACATTGCTAGTTGAATCTGCTGCTAATACATTAGCTTCA 3513
              |||::: ||| ::: ||| |||::: |||
QY      424  ValAlaAspGlnThrValLysLeuAlaLysThrValLysAspMetSerGlyGlu----- 441
              ||| ||| ||| ||| |||
Db      3514  GTAGCTGACCAAGCTGCAATCTTGCTGCTACAAATTGATACTTTAAACACTGCAGATAAG 3573
              ||| ||| ||| ||| |||
QY      442  ---AlaIleSerSerThrGlyAla-----SerLeuArgSerThrVal 454
              |||::: ||| ||| ||| |||
Db      3574  TTTCACAGCTTCTGCACACAGGTGCTACTATTACATTAACTTCTACTGTA 3621
              |||::: ||| ||| ||| |||

RESULT 13
US-08-682-517-7
; Sequence 7, Application US/08682517
; Patent No. 5874267
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Expression of surface layer proteins
; NUMBER OF SEQUENCES: 25
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/682,517
; FILING DATE:
; CLASSIFICATION:
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4197 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single

```

[illegible]

QY 224 ArgThrValLeuAlaProAlaLeuAlaSerArgProSerValGlnGlyAlaValAspPhe 243
Db 3065 ---ACTGTTCAAGCTGGTACATTATCTTCT----- 3091
QY 244 GlyValSerThrAlaGlyGlyLeuValAlaAsn-----AlaGly 256
Db 3092 -----TTAACTGCTGGTACATTAAACAGTACTTATGCAGATGCTAAATAATGCTGCAGGT 3145
QY 257 PheGlyAspArgMetLeuSerValGlnSerArgAspGlnLeuArgGlyGly----- 273
Db 3146 GTTGCTGAAAAATATTACTGCTAGCGTAACATTAAAGAAACTACTGGAGCAATTACTTCT 3205
QY 274 ---AlaPheValLeuGlyMetLysAspLysGlnProLysAlaAlaLeuSerGlnGluThr 292
Db 3206 GATACATTATACACAGGTGTATTA-----CCATCAGCAGCTACAGCAGCTGAATAT 3256
QY 293 AspTrpLeuAspAlaTyrLysAlaIleLysSerAlaSerTyrSerGlyAlaAlaLeu--- 311
Db 3257 ACTCTAAATCAATTGCTGCAGATTATACATTTCACACAGGTGAAGGATTCACCTTTAAAT 3316
QY 312 -----AsnAlaGlyLysArgMetAlaGlyLeuProLeuAspValAlaThrAspGlyLeu 329
Db 3317 ATGTATTAATGCTGGTGCCTCAAGTAATTAACCTTAGCA-----GGTAAA 3358
QY 330 LysAlaValArgSerLeuValSerAlaThrSerLeuThrLysAsnGlyLeuAlaLeuAla 349
Db 3359 AAAGTGCACAGAGTGTAGCTGATGCTATCAATGCTACATTTCAGAGTACTGCAACTGTT 3418
QY 350 GlyGlyTyrAlaGlyValSerLysLeuGlnLysMetAlaThrLysAsnIleThrAsp--- 368
Db 3419 TCTGGA-----GACAAAGTAGTTATTAAATCAGCTACACACAGGT 3457
QY 369 -----SerAlaThrLysAlaAlaValSerGlnLeuSerAsnLeuVal--- 382
Db 3458 GTTGGTCTGAGTTGAAGTTACATTCTCTTCTGTTAATCAAGTATTAAATGACAGTAGTT 3517
QY 383 ---GlySerValGlyValPheAlaGlyTyr---ThrThrAlaGlyLeuAlaThrAspPro 400
Db 3518 AACGGTAAAGATCAAGTCGTTGCAGGAACAGCTGCTACAAAAGCATTCACGATTACTACA 3577
QY 401 AlaValLysLysAlaGlyLysPheIleGlnAspLysValLysSerThr----- 416
Db 3578 GCCCTTCTGTTGGGTGAAAAAGTAGTTATGATGTTGTAATATACTGCTGATGCAATTT 3637
QY 417 -----AlaSerSerThrThrSerTyr 423
Db 3638 GGAAGTGTCCACACAGCAATACATTGCTAGTTGAATCTGCTGAATACATTAGCTTCA 3697
QY 424 ValAlaAspGlnThrValLysLeuAlaLysThrValLysAspMetSerGlyGlu----- 441
Db 3698 GTAGCTGACCAAGCTGCAAAATCTTGTCTGCTACAAATTGATACTTTAAACACTGCAGATAAG 3757
QY 442 ---AlaIleSerSerThrGlyAla-----SerLeuArgSerThrVal 454
Db 3758 TTACAGCTTCTGCAACACAGGTGCTACTATTACATTACTTCTACTGTA 3805

RESULT 14
US-08-682-517-8
; Sequence 8, Application US/08682517
; Patent No. 5874267
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Expression of surface layer proteins
; NUMBER OF SEQUENCES: 25
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/682,517
; FILING DATE:

CLASSIFICATION:
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4197 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Bacillus sphaericus
; INDIVIDUAL ISOLATE: P-1
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 95..3850
; FEATURE:
; NAME/KEY: mat_peptide
; LOCATION: 185..3850
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 95..184
US-08-682-517-8

Alignment Scores:
Pred. No.: 0.00446 Length: 4197
Score: 125.50 Matches: 117
Percent Similarity: 32.99% Conservative: 73
Best Local Similarity: 20.31% Mismatches: 219
Query Match: 5.21% Indels: 167
DB: 2 Gaps: 25

US-09-825-414-66 (1-487) x US-08-682-517-8 (1-4197)

QY 3 IleAsnGlnSerAlaGlnGlnProProGlyValAlaMetGlu----- 16
Db 2207 GTAACCAAGCGCAAAAAAGTAACCTGGTACTTCTAATTAAGAAAGCACTTATACAATC 2266
QY 17 -----Ser 17
Db 2267 TACATACTGGTGTAAATGATATTAAAGTAGATTAAACCAAGTTATCTCACCAAAATCGTAGC 2326
QY 18 PheArgThrAlaSerAspAlaSerLeuAlaSerSerSerValArgSerValSerThrThr 37
Db 2327 TACACAGTAACTTATGAAGCTACTTATCTTCTACA-----GGAAGCTGTTATTACA 2377
QY 38 SerCysArgAspLeuGlnAlaIleThrAspTyrLeuLysHisHisValPheAlaAlaHis 57
Db 2378 CCTGCTAAGAATTAGAAAGTTACTTTCAGTGGATGGTTAAACCAACTGCTGTTAAAGTAATT 2437
QY 58 ArgPheSerValIleGlySerProAspGluArgAspAlaAlaLeuAlaHisAsnGluGln 77
Db 2438 GCTACAGGTATGCTGTTAATACAGACGGTAAAGACTATGCATTACTGCTAAAGAAAGCT 2497
QY 78 IleAspAlaLeuValGluThr----- 84
Db 2498 ACAGCTACATTTCACAGCTACAAATGAAGTTCCAAACCTTTACACACTGGTGTAGCTACTCAA 2557
QY 85 -----ArgAlaAsnArgLeuTyrSerGluGlyGluThrPro 96
Db 2558 TTTCATACAGCTGAATTCTGTTCAAAACACAGCACTCTATTGTTGCTGGTAAAAACCCA 2617
QY 97 AlaThrIleAlaGluThrPheAlaLysAlaGlyLysPheAspArgLeuAlaThrThrAla 116
Db 2618 GTGAATATGCTGGTGTATCAGGCAAAACATTAATAATAC-----TTCCGAGCTAATGGT 2671
QY 117 SerSerAlaPheGluAsnThrProPheAlaAlaAlaSerValLeuGlnTyr----- 133
Db 2672 AATGAAGTATTTGGTGAAGCGGCATGGGAGCATTTA---TTAACTCAATATGCAACTGAA 2728
QY 134 -----MetGln 135
Db 2729 GGCCAAAAAGTAAACAATCTCATATATAGTAGATGCTATACAGTTACATTAAAGTAATT 2788
QY 136 ProAlaIleAsnLysGlyAspTrpLeuAlaThrProLeuLysProLeuThrProLeu--- 154

Db 2789 AGTGGCTGTTAATTCTTCAACTGGAAGCTATCAACCAAGTTGGTCTCCACACACAGCAGCT 2848
|||:::||||
QY 155 ---IleSerGlyAlaLeuSer-----GlyAlaMetasp----- 164
:::|||||:::
Db 2849 CCAACTACTGGCGCATTAACATTAAACACAGCAGCTGGTGGTTGATTGATTAAACAAC 2908
|||
QY 165 GlnValGlyThrLysMetMetaspArgAlaArgGlyAspLeuHisTyrLeuSerThrSer 184
|||
Db 2909 GCAACTAACACTTTAGGAATTTTCATTAGCTGATGCAGATCTTAATGTAGTGCACAAC 2968
|||
QY 185 ProAspLysLeuHisAspAlaMetAlaValSerValLys---ArgHisSerProAlaLeu 203
|||
Db 2969 GTT-----GATACTGCAACTGTTTCATTAAAGATAGTCAAAATTAATTCATT 3016
|||
QY 204 GlyArgGlnValAlaValAspMetGlyIleAlaValGlnThrPheSerAlaLeuAsnValVal 223
:::||||:::
Db 3017 TCTCTTACATTAGTTGAACCTGGTCTTAATACAGGTGATTGTTGCTACA----- 3064
|||
QY 224 ArgThrValLeuAlaProAlaLeuAlaSerArgProSerValGlnGlyAlaValAspPhe 243
|||||
Db 3065 ---ACTGTTCAAGCTGGTACATTATCTTCT----- 3091
|||
QY 244 GlyValSerThrAlaGlyGlyLeuValAlaAsn-----AlaGly 256
|||||
Db 3092 -----TTAACTGCTGGTACATTAACTTACTTATGACAGATGCTAAAAATGCTGACAG 3145
|||
QY 257 PheGlyAspArgMetLeuSerValGlnSerArgAspGlnLeuArgGlyGly----- 273
:::|:::
Db 3146 GTTGGTGAATAATATTACTGCTAGCGTTACATTAAAGAAACTACTGAGCAATTACTTCT 3205
|||
QY 274 ---AlaPheValLeuGlyMetLysAspLysGluProLysAlaAlaLeuSerGluGluThr 292
|||
Db 3206 GATACATTTACACAAGGTGTATT-----CCATCAGCAGCTACAGCAGCTGAATAT 3256
|||
QY 293 AspTrpLeuAspAlaTyrLysAlaIleLysSerAlaSerTyrSerGlyAlaAlaLeu--- 311
|||
Db 3257 ACTTCTAAATCAATTGCTGCAGATTATACATTGTCACACAGGTGAAGATTCACTTAAAT 3316
|||
QY 312 -----AsnAlaGlyLysArgMetAlaGlyLeuProLeuAspValAlaThrAspGlyLeu 329
|||||
Db 3317 ATTGATTAATGCTGGTGTCTCAAGTAAATTAACTTAGCA-----GTTAAA 3358
|||
QY 330 LysAlaValArgSerLeuValSerAlaThrSerLeuThrLysAsnGlyLeuAlaLeuAla 349
|||
Db 3359 AAAGGTGCACAAGGTGTAGCTGATGCTATCAATGCTACATTGTCAGTACTGCAACTGTT 3418
|||
QY 350 GlyGlyTyrAlaGlyValSerLysLeuGlnLysMetAlaThrLysAsnIleThrAsp--- 368
|||
Db 3419 TCTGGA-----GACAAAGTAGTTATTAATCAGCTACACACAGGT 3457
|||
QY 369 -----SerAlaThrLysAlaAlaValSerGlnLeuSerAsnLeuVal--- 382
|||
Db 3458 GTTGGTTCTGAAGTTGAAGTTACATTCTCTTCTGTTAATCAAGTATTAAATGCAGTAGTT 3517
|||
QY 383 ---GlySerValGlyValPheAlaGlyTyr---ThrThrAlaGlyLeuAlaThrAspPro 400
|||
Db 3518 AACGGTAAAGATCAAGTCGTTGCGAAGACAGCTGCTACAAAGCATTCACGATTACTACA 3577
|||
QY 401 AlaValLysLysAlaGlyLeuSerPheIleGlnAspLysValLysSerThr----- 416
|||
Db 3578 GCCCTTCTGTGGTGAAAAAGTAGTTATTGATGGTGAATATTAATCTGCTGACATTT 3637
|||
QY 417 -----AlaSerSerThrThrSerTyr 423
|||
Db 3638 GGAACCTGCTCCACACAGCAAAATACATTGCTAGTTGAATCTGCTGCTAATACATTAGCTTCA 3697
|||
QY 424 ValAlaAspGlnThrValLysLeuAlaLysThrValLysAspMetSerGlyGlu----- 441
|||
Db 3698 GTAGCTGACCAAGCTGCAAAATCTTCTGCTACAAATTGATTAATAACACTGCAGATTAAG 3757
|||
QY 442 ---AlaIleSerSerThrGlyAla-----SerLeuArgSerThrVal 454
|||

Db 3758 TTACAGCTTCTGCAACAGGTGCTACTATTACATTAACTTCTACTGTA 3805
RESULT 15
US-09-141-047-7
; Sequence 7, Application US/09141047A
; Patent No. 6043085
; GENERAL INFORMATION:
; APPLICANT: Walker, David H.
; APPLICANT: Yu, Xue-Jie
; TITLE OF INVENTION: Ehrlichia canis 120-kDa Immunodominant Antigenic
; TITLE OF INVENTION: Protein Gene
; FILE REFERENCE: D6143
; CURRENT APPLICATION NUMBER: US/09/141,047A
; CURRENT FILING DATE: 1998-08-27
; NUMBER OF SEQ ID NOS: 15
; SEQ ID NO 7
; LENGTH: 2489
; TYPE: DNA
; ORGANISM: Ehrlichia canis
; FEATURE:
; LOCATION: -340..2149
; OTHER INFORMATION: Nucleotide sequence of gene encoding 120kDa
; OTHER INFORMATION: Immunoreactive protein
US-09-141-047-7

Alignment Scores:
Pred. No.: 0.00351 length: 2489
Score: 123.00 Matches: 116
Percent Similarity: 35.69% Conservative: 76
Best Local Similarity: 21.56% Mismatches: 208
Query Match: 5.11% Indels: 138
DB: 3 Gaps: 26
US-09-825-414-66 (1-487) x US-09-141-047-7 (1-2489)

QY 8 GlnGlnProGlyValAlaMetGluSerPheArgThrAlaSerAspAlaSerLeuAla 27
:::|:::
Db 653 GAAGTACTCTCTGAAGTTAAAGCAGAAAGATTGCAACCTGCTGATGATGATAGAA 712
|||
QY 28 SerSerSerVal-----ArgSerValSerThrThrSer----- 38
|||||
Db 713 CATTCATCAAGTGAAGTTGGAGAAAAGTATCTAAACTAGTAAGAGAAAGTACTCT 772
|||
QY 39 -----CysArgAspLeuGln---AlaIleThrAspTyrLeuLysHisHisValPhe 54
|||
Db 773 GAAGTTAAAGCAGAAAGATTGCAACCTGCTGATGATGATGATGATGATGATGATGAT 820
|||
QY 55 AlaAlaHisArgPheSerValIleGly----- 63
|||
Db 821 -----CATTCATCAAGTGAAGTTGGAGAAAAGTATCTGAAACTAGTAAAGAGCAAAAT 874
|||
QY 64 SerProAspGluArg-----AspAlaAlaLeuAlaHisAsn 75
:::|:::
Db 875 ACTGCTGAAGTTAAAGCAGAAAGATTGCAACCTGCTGATGATGATGATGATGATGATGAT 934
|||
QY 76 GluGlnIleAspAlaLeuValGluThrArgAlaAsnArgLeuTyrSerGluGlyGluThr 95
:::|:::
Db 935 -----TCAAGTGAAGTTGGAGAAAAGTATCTAAACTAGTAAAGAGAAAGTACT 985
|||
QY 96 ProAlaThrIleAlaGluThrPhe-----AlaLys 105
|||
Db 986 CCTGAAGTTAAAGCAGAAAGATTGCAACCTGCTGATGATGATGATGATGATGATGATGAT 1045
|||
QY 106 AlaGluLysPheAspArgLeuAlaThrThrAlaSerSerAlaPheGluAsnThrProPhe 125
:::|:::
Db 1046 AGTGAAGTTGGAGAAAAGTATCTGAAACTAGTAAAGAA-----GAAATATCTCTGAA 1099
|||
QY 126 AlaAlaAlaSerValLeuGlnTyrMetGlnProAlaIleAsnLysGlyAspTrpLeuAla 145
|||
Db 1100 GTTAAACCGGAAGATTGGAA----- 1120
|||
QY 146 ThrProLeuLysProLeuThrProLeuIleSerGlyAlaLeuSerGlyAlaMetaspGln 165
|||

Db 1121 -----CCTGCTGTAGATGCTAGTGTAGAACATTCATCAAGTGAA 1159

QY 166 ValGlyThrLysMetMetAsparGlyAlaArgLysPLeuHisTyrLeuSerThrSerPro 185
||||| : : : : : |||||

Db 1160 GTTGGGAAAAAGTATCTGAACTAGTAAGAGGAA-----AGTACTCCTGAA 1207

QY 186 AsplysLeuHisasp-----AlaMetAlaValSerValLysArgHisSerProAla 202
||| : : : : : |||||

Db 1208 GTTAAAGCAGAGATTTGCCAACCCTGCTGATGATGATGTGGACATTCATCAAGTGAA 1267

QY 203 LeuGlyArgGlnValValAspMetGlyIleAlaValGlnThrPheSerAlaLeuAsnVal 222
:::| | : : : : : |||

Db 1268 GTTGAGAAAAAGTATCTGAAACTAGTAAGAGGAAAAATACT----- 1309

QY 223 ValArgThrValLeuAlaProAlaLeuAlaSerArgProSerValGlnGlyAlaValAsp 242
||| : : : : : ||| : : : : : ||| : : : : : |||

Db 1310 ---CCTGAAGTTAGACAGACAGATTTG-----CAACCTGCTGATGATGCTAGTAGAA 1360

QY 243 PheGlyValSerThrAlaGlyGlyLeuValAlaAsnAlaGlyPheGlyAspArgMetLeu 262
||| : : : : : ||| : : : : : |||

Db 1361 CATTCATCAAGTGAAGTTGGAGAAAAAGTATCTGAAACTAGTAAGAGGAAAGTACTCCT 1420

QY 263 SerValGlnSerArgasp-----GlnLeuArgGlyGlyAlaPhe 275
||| : : : : : ||| : : : : : |||

Db 1421 GAAGTTAAAGCAGAGATTTGCCAACCCTGCTGATGATGATGATGATGATGATGATGAT 1480

QY 276 ValLeuGlyMetLysAspLysGlu---ProLysAlaAlaLeuSerGluIuThrAspTrp 294
:::| | : : : : : ||| : : : : : |||

Db 1481 GAAGTTGGGAAAAAAGTATCTGAAACTAGTAAGAGGAAAGTACTCCTGAAAGTTAAAGCA 1540

QY 295 LeuAspAlaTyrLysAlaIle-----LysSerAlaSerTyrSerGlyAla 309
||| : : : : : ||| : : : : : |||

Db 1541 GAAGATTTGCCAACCCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGAA 1600

QY 310 AlaLeuAsnAlaGlyLysArgMetAlaGlyLeuPro-----LeuAsp 323
:::| | : : : : : |||

Db 1601 AAAGTATCTGAAACTAGTAAGAGGAAAAAAGTACTCCTGAAAGTTAAAGCAGAGATTTGCAA 1660

QY 324 ValAlaThrAspGly-----LeuLysAlaValArgSerLeuValSerAla 338
||| : : : : : ||| : : : : : |||

Db 1661 CCTGCTGTAGATGCTAGTGTAGAACATTCATCAAGTGAAGTTGGAGAAAAAGTATCTGAA 1720

QY 339 ThrSerLeuThrLysAsnGlyLeuAlaLeuAlaGlyGlyTyrAlaGlyValSerLysLeu 358
||||| : : : : : ||| : : : : : |||

Db 1721 ACTAGTAAGAGGAAAAATACTCCTGAGATTAA-----GCCGAGATTTTG 1765

QY 359 GlnLysMetAlaThrLysAsnIleThrAspSerAlaThrLysAlaAla-----ValSer 376
||| : : : : : ||| : : : : : |||

Db 1766 CAACCTGCTGTAGATGCTAGTGTAGAACATTCATCAAGTGAAGTTGGAGAAAAAGTATCT 1825

QY 377 GlnLeuSerAsnLeuValGlySerValGlyValPheAlaGlyTyrThrAlaGlyLeu 396
:::| | : : : : : ||| : : : : : |||

Db 1826 GAAACTAGTAAGAGGAAAGTACTCCTGAGTTAAAGCG-----GAAGATTTGCCAACCT 1879

QY 397 AlaThrAspProAlaValLysLysAlaGlySerPheIleGlnAspLysValLysSerThr 416
||| : : : : : ||| : : : : : |||

Db 1880 GCTGTAGATGCTAGTGTGGAACATTCATCAAGTGAAGTTGGAGAAAAAGTATCTGAGACT 1939

QY 417 Ala-----SerSerThrThrSerTyrValAla---AspGlnThrValLysLeuAlaLys 433
:::| | : : : : : ||| : : : : : |||

Db 1940 AGTAAGAGAGAAAGTACTCCTGAGATTAAAGCGGAGAGATTTGCCAACCTGCTGATGATGCT 1999

QY 434 ThrValLysAspMetSer-----GlyGlnAlaIleSerSerThrGlyAlaSerLeu 450
:::| | : : : : : ||| : : : : : |||

Db 2000 AGTGTGAACATTCATCAAGTGAAGTTGGAGAAAAAGTATCTGAGACTAGT----- 2050

QY 451 ArgSerThrValAsnAsnLeuArgHisArgSerAlaProGlnAlaAspIleGlu 468
:::| | : : : : : ||| : : : : : |||

Db 2051 -----AAAGAGGAAAGTACTCCTGAAAGTTAAAGCGGAA 2083

GenCore version 5.1.3
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - nucleic search, using frame_plus.p2n model

Run on: January 31, 2003, 04:32:35 ; Search time 77.0791 Seconds
(without alignments)
2838.537 Million cell updates/sec

Title: US-09-825-414-66

Perfect score: 2407

Sequence: 1 MHINSAQPPGVAMESFRT.....EEGGISAFSRSETPFQLRRL 487

Scoring table:

BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 396772 segs, 224632407 residues

Total number of hits satisfying chosen parameters: 793544

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:

-MODEL=frame+_p2n.model -DEV=xlp
-Q=/cgn2_1/USPTO_spool/US09825414/runat_28012003_161742_26139/app_query.fasta_1.1294
-DB=Published_Applications_NA -QFMT=fastap -SUFFIX=p2n.rnpb -MINMATCH=0.1
-LOOPCL=0 -LOOEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62
-TRANS=human40.cdl -LIST=45 -DOCCALIGN=200 -THR_SCORE=pct -THR_MAX=100
-THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0
-MAXLEN=200000000 -USER=US09825414_@CGN_1_1_46_@runat_28012003_161742_26139
-NCPU=6 -ICPU=3 -NO_XLPXY -NO_MAP -LARGEQUERY -NEG_SCORES=0 -WAIT -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published_Applications_NA:*

1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2407	100.0	1464	10	US-09-825-414-65 Sequence 65, Appli
2	1869.5	77.7	1461	10	US-09-825-414-6 Sequence 6, Appli
3	1869.5	77.7	30365	10	US-09-825-414-1 Sequence 1, Appli
4	147	6.1	4645	9	US-09-971-536-27 Sequence 27, Appli

5	146.5	6.1	9542	9	US-09-813-214A-8	Sequence 8, Appli
6	140	5.8	7100	10	US-09-932-183A-1	Sequence 1, Appli
7	138	5.7	8979	9	US-09-738-626-2739	Sequence 2739, Ap
8	132	5.5	3300	10	US-09-379-931-6	Sequence 6, Appli
9	130.5	5.4	7104	10	US-09-815-242-4580	Sequence 4580, Ap
10	130.5	5.4	7107	10	US-09-815-242-8291	Sequence 8291, Ap
11	129.5	5.4	6228	10	US-09-815-242-4760	Sequence 4760, Ap
12	129.5	5.4	6561	10	US-09-815-242-8815	Sequence 8815, Ap
13	129	5.4	7035	10	US-09-815-242-8615	Sequence 8615, Ap
14	127.5	5.3	7434	10	US-09-815-242-4761	Sequence 4761, Ap
15	127.5	5.3	7437	10	US-09-815-242-8869	Sequence 8869, Ap
16	127	5.3	3309400	9	US-09-738-626-1	Sequence 1, Appli
17	125.5	5.2	3666	10	US-09-137-531-13	Sequence 13, Appli
18	125.5	5.2	3666	10	US-09-137-531-14	Sequence 14, Appli
19	125.5	5.2	4197	10	US-09-137-531-7	Sequence 7, Appli
20	125.5	5.2	4197	10	US-09-137-531-8	Sequence 8, Appli
21	125	5.2	6258	10	US-09-815-242-4584	Sequence 4584, Ap
22	125	5.2	17388	10	US-09-815-242-8512	Sequence 8512, Ap
23	124	5.2	1887	10	US-09-841-786-12	Sequence 12, Appli
24	124	5.2	9726	10	US-09-841-786-8	Sequence 8, Appli
25	124	5.2	11130	10	US-09-841-786-15	Sequence 15, Appli
26	123.5	5.1	7302	10	US-09-815-242-4780	Sequence 4780, Ap
27	123.5	5.1	18846	10	US-09-815-242-8898	Sequence 8898, Ap
28	123	5.1	3006	9	US-10-124-800-16	Sequence 16, Appli
29	123	5.1	8730	9	US-10-124-800-1	Sequence 1, Appli
30	123	5.1	11458	10	US-09-825-414-18	Sequence 18, Appli
31	120.5	5.0	1734	10	US-09-974-300-1874	Sequence 1874, Ap
32	120.5	5.0	2433	9	US-09-712-363-137	Sequence 137, App
33	119.5	5.0	2337	10	US-09-815-242-6631	Sequence 6631, Ap
34	119.5	5.0	3087	10	US-09-815-242-4830	Sequence 4830, Ap
35	119.5	5.0	3144	10	US-09-815-242-8985	Sequence 8985, Ap
36	119	4.9	1801	10	US-09-287-849-21	Sequence 21, Appli
37	118.5	4.9	3832	9	US-10-147-026-7	Sequence 7, Appli
38	118	4.9	1941	10	US-09-841-132-412	Sequence 412, App
39	118	4.9	2436	9	US-10-063-547-99	Sequence 99, Appli
40	118	4.9	2436	9	US-10-174-590-309	Sequence 309, App
41	118	4.9	2436	9	US-10-176-758-309	Sequence 309, App
42	118	4.9	2436	9	US-10-063-616-99	Sequence 99, Appli
43	118	4.9	2436	9	US-10-175-737-309	Sequence 309, App
44	118	4.9	2436	12	US-10-006-867-99	Sequence 99, Appli
45	118	4.9	2436	12	US-10-052-586-309	Sequence 309, App

ALIGNMENTS

RESULT 1
US-09-825-414-65
; Sequence 65, Application US/09825414
; Patent NO. US20020083489A1
; GENERAL INFORMATION:
; APPLICANT: Colimer, Alan
; APPLICANT: Alfano, James R.
; APPLICANT: Charkowski, Amy O.
; TITLE OF INVENTION: DNA MOLECULES AND POLYPEPTIDES OF PSEUDOMONAS SYRINGAE
; FILE REFERENCE: 19603/3243
; CURRENT APPLICATION NUMBER: US/09/825,414
; PRIOR FILING DATE: 2001-04-03
; PRIOR APPLICATION NUMBER: 60/194,160
; PRIOR FILING DATE: 2000-04-03
; PRIOR APPLICATION NUMBER: 60/224,604
; PRIOR FILING DATE: 2000-08-11
; PRIOR APPLICATION NUMBER: 60/249,548
; PRIOR FILING DATE: 2000-11-17
; NUMBER OF SEQ ID NOS: 91
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 65
; LENGTH: 1464
; TYPE: DNA
; ORGANISM: Pseudomonas syringae pv. tomato
US-09-825-414-65
Alignment Scores:

Pred. No.: 6.63e-240 Length: 1464
Score: 2407.00 Matches: 487
Percent Similarity: 100.00% Conservative: 0
Best Local Similarity: 100.00% Mismatches: 0
Query Match: 100.00% Indels: 0
DB: 10 Gaps: 0

US-09-825-414-66 (1-487) x US-09-825-414-65 (1-1464)

QY 1 MethisIleasnGlnSerAlaGlnGlnProGlyValAlaMetGluSerPheArgThr 20
Db 1 ATGCACATCAACCAATCCGCCCAACAACCGCTGGCTTGCAATGGAGAGTTTCGGACA 60
QY 21 AlaSerAspAlaSerLeuAlaSerSerSerValArgSerValSerThrThrSercysArg 40
Db 61 GCTTCGACGCGCTCCCTGCTTCGAGTTCTGTGGGTTCTGCAGCACTACCTCGTCCGC 120
QY 41 AspLeuGlnAlaIleThrAspTyrLeuLysHisHisValPheAlaAlaHisArgPheSer 60
Db 121 GATCTACAAGCTATTACGATTATCTGAACAACATACGTGTCTGCTGCACAGGTTTCG 180
QY 61 ValIleGlySerProAspGluArgAspAlaAlaLeuAlaHisasnGluGlnIleAspAla 80
Db 181 GTAATAGGCTACACCGGATGAGCGTGATCCGCTTGACACACAGACAGATCGATCGC 240
QY 81 LeuValGluThrArgAlaasnArgLeuTyrSerGluGlyGluThrProAlaThrIleAla 100
Db 241 TTGGTAGAGACACGCGCAACCGCCTGTACTCCGAGGGAGACCCCGCAACCATCCGC 300
QY 101 GluThrPheAlaLysAlaGluLysPheAspArgLeuAlaThrThrAlaSerSerAlaPhe 120
Db 301 GAACATTCGCCAAGCGCAAAAGTTCGACCCGTTGGCGACGACCGCATCAAGTCCTTT 360
QY 121 GluAsnThrProPheAlaAlaAlaSerValLeuGlnTyrMetGlnProAlaIleasnLys 140
Db 361 GAGAACACGCCCATTTGCCGCTGCCCTCGTTCAGTACATGCAGCCCTCGATCAACAAG 420
QY 141 GlyAspTrpLeuAlaThrProLeuLysProLeuThrProLeuIleSerGlyAlaLeuSer 160
Db 421 GCGGATTGGCTAGCAACGCCGCTCAAGCCGCTGACCCCGCTCATTTCCGAGCGGTGCG 480
QY 161 GlyAlaMetAspGlnValGlyThrLysMetMetAspArgAlaArgGlyAspLeuHisTyr 180
Db 481 GGAGCCATGGACCAAGTGGGCACCAAAATGATGATCGTGCAGGGGTGATCGATTAC 540
QY 181 LeuSerThrSerProAspLysLeuHisAspAlaMetAlaValSerValLysArgHisSer 200
Db 541 CTGAGCACTTCGCGGACAAGTTGCATGATGGCGGCTATCGGTGAAGCGCACTCG 600
QY 201 ProAlaLeuGlyArgGlnValValAspMetGlyIleAlaValGlnThrPheSerAlaLeu 220
Db 601 CCTGCCGCTGGTCGACAGGTGTGACATGGGGATTCAGTGCAGACGTTCTCGCGCTA 660
QY 221 AsnValValArgThrValLeuAlaProAlaLeuAlaSerArgProSerValGlnGlyAla 240
Db 661 AATGTGGTGCATTGCTATTGGCTCCAGCACCTAGCGTCCAGACCGTGGTGCAGGGTCT 720
QY 241 ValAspPheGlyValSerThrAlaGlyGlyLeuValAlaasnAlaGlyPheGlyAspArg 260
Db 721 GTTGATTTGGCGTATCTACGCGGGGTGGCTTGGTTCGAATGCAGGCTTTGGCGACCGC 780
QY 261 MetLeuSerValGlnSerArgAspGlnLeuArgGlyGlyAlaPheValLeuGlyMetLys 280
Db 781 ATGCTCAGTGTCAATCGCGCGATCAACTCGGTGGGGGGGCATTCGTACTTGGCATGAAA 840
QY 281 AspLysGluProLysAlaAlaLeuSerGluGluThrAspTrpLeuAspAlaTyrLysAla 300
Db 841 GATTAAGAGAGCCCAAGCGCGCTTGAGTGAAGAACTGATTGGCTTGATGCTTACAAAGCG 900
QY 301 IleLysSerAlaSerTyrSerGlyAlaAlaLeuAsnAlaGlyLysArgMetAlaGlyLeu 320
Db 901 ATCAAGTCGGCCAGCTACTCAGGTGCGGCGCTCAATGCGGCAAGCGGATGGCGCGCTG 960

QY 321 ProLeuAspValAlaThrAspGlyLeuLysAlaValArgSerLeuValSerAlaThrSer 340
Db 961 CCACGTGACGTGCGGACCGACCGGCTCAAGGCGGCTGAGAAGTCTGTGTGCGCCACCAGC 1020
QY 341 LeuThrLysAsnGlyLeuAlaLeuAlaGlyGlyTyrAlaGlyValSerLysLeuGlnLys 360
Db 1021 CTGACAAAAAATGGCTGGCCCTAGCCGCTGGTTACGCGGGGTAACTAAGTTGCAGAAA 1080
QY 361 MetaIaThrLysAsnIleThrAspSerAlaThrLysAlaAlaValSerGlnLeuSerAsn 380
Db 1081 ATGGCGACGAAAAATATCACTGATTCGGCGACCAAGGCTCGGTAGTCACTGACGACAAC 1140
QY 381 LeuValGlySerValGlyValPheAlaGlyTyrThrThrAlaGlyLeuAlaThrAspPro 400
Db 1141 CTGTTGGGTTCGGTAGCGCTTTTCGACAGGCTGGACCAACCGCTGGACTGGCAGTACCT 1200
QY 401 AlaValLysLysAlaGluSerPheIleGlnAspLysValLysSerThrAlaSerSerThr 420
Db 1201 GCGGTTAAGAAAGCCGAGTCTGTTATACAGAGATTAAGGTGAATCGACCGCATCTAGTACC 1260
QY 421 ThrSerTyrValAlaAspGlnThrValLysLeuAlaLysThrValLysAspMetSergly 440
Db 1261 ACAAGCTATGTTGCCGACCAAGACCGTCAAACTGCGGAAACAGTCAAGGCATGAGCGGG 1320
QY 441 GluAlaIleSerSerThrGlyAlaSerLeuArgSerThrValAsnAsnLeuArgHisArg 460
Db 1321 GAGCGATCTCCAGCACCGGCTGCCAGCTTACGCACTACTGTCAATAACTGCGTCATCGC 1380
QY 461 SerAlaProGluAlaAspIleGluGlyGlyLysSerAlaPheSerArgSergluThr 480
Db 1381 TCCGCTCCGGAAGCTGATATCGAAGAAAGTGGGATTTCCGCGTTTCTCGAAGTGAACA 1440
QY 481 ProPheGlnLeuArgArgLeu 487
Db 1441 CCGTTTCAGCTCAGCGCTTTG 1461

RESULT 2

US-09-825-414-6

; Sequence 6, Application US/09825414

; Patent No. US20020083489A1

; GENERAL INFORMATION:

; APPLICANT: Colimer, Alan

; APPLICANT: Alfano, James R.

; APPLICANT: Charkowski, Amy O.

; TITLE OF INVENTION: DNA MOLECULES AND POLYPEPTIDES OF PSEUDOMONAS SYRINGAE

; TITLE OF INVENTION: HRP PATHOGENICITY ISLAND AND THEIR USES

; FILE REFERENCE: 19603/3243

; CURRENT APPLICATION NUMBER: US/09/825,414

; PRIOR APPLICATION NUMBER: 60/194,160

; PRIOR FILING DATE: 2000-04-03

; PRIOR APPLICATION NUMBER: 60/224,604

; PRIOR FILING DATE: 2000-08-11

; PRIOR APPLICATION NUMBER: 60/249,548

; PRIOR FILING DATE: 2000-11-17

; NUMBER OF SEQ ID NOS: 91

; SOFTWARE: Patentin Ver. 2.1

; SEQ ID NO 6

; LENGTH: 1461

; TYPE: DNA

; ORGANISM: Pseudomonas syringae

; US-09-825-414-6

Alignment Scores:

Pred. No.: 2.67e-184 Length: 1461
Score: 1869.50 Matches: 381
Percent Similarity: 85.19% Conservative: 33
Best Local Similarity: 78.40% Mismatches: 71
Query Match: 77.67% Indels: 1
DB: 10 Gaps: 1

US-09-825-414-66 (1-487) x US-09-825-414-6 (1-1461)

QY	1	MethisIleasnGlnSerAlaGlnGlnProProGlyValAlaMetGluSerPheArgThr	20
Db	1	ATGCACATCAACCGACGCGTCCAAACACCGCCTGTGACTGCGACGGATAGCTTTCGACA	60
QY	21	AlaSerAspAlaSerLeuAlaSerSerSerValArgSerValSerThrThrSerCysArg	40
Db	61	GGCTCCGACGCGGTCTTGTCCCTCCAGCTCTGTGGCATCTGTACAGCTCCGATCAGCAACGC	120
QY	41	AspLeuGlnAlaIleThrAspTyrLeuLysHisHisValPheAlaAlaHisArgPheSer	60
Db	121	GAGATTAATGCGATTGCCGATTACCTGACACAGATCATGTGTTCGCTGCCATAA	180
QY	61	ValIleGlySerProAspGluArgAspAlaAlaLeuAlaHisAsnGluGlnIleAspAla	80
Db	181	CCGGCCGATTGGCTGATGGCCAAAGCTGCAGTTGACATACAAATGCGCAGATCACTGCC	240
QY	81	LeuValGluThrArgAlaAsnArgLeuTyrSerGluGlyGluThrProAlaThrIleAla	100
Db	241	CTGATCGAGACGCGCCGACGCGCTGCACCTCGAAGGGGAAACCCCGCAACCATCGCC	300
QY	101	GluThrPheAlaLeuAlaGluLysPheAspArgLeuAlaThrThrAlaSerSerAlaPhe	120
Db	301	GACACCTTCGCCAAGCGGAAAGCTCGACCGATTGGCGACGACTACATCAGGCCCGTTG	360
QY	121	GluAsnThrProPheAlaAlaAlaSerValLeuGlnTyrMetGlnProAlaIleAsnLys	140
Db	361	CGGCGACGCGCCTTGGCCATGGCCTCGTTGCTTCACTACATGCACGCTGCATCAACA	420
QY	141	GlyAspTrpLeuAlaThrProLeuLysProLeuThrProLeuIleSerGlyAlaLeuSer	160
Db	421	GGCGATTGGCTGGCGGCTCCGCTCAACCGCTGACCCCGCTCATTTCCGAGCGCTGCG	480
QY	161	GlyAlaMetaspGlnValGlyThrLysMetMetAspArgAlaArgGlyAspLeuHisTyr	180
Db	481	GGCGCCATGGACCAAGTGGCCACCAAGATGATGACCGCGGAGCGGTGATCTGCATTAC	540
QY	181	LeuSerThrSerProAspLysLeuHisAspAlaMetAlaValSerValLysArgHisSer	200
Db	541	CTGAGCGCCTGCGCGACAGGCTCCACGATGCGATGGCGGCTTGGTGAAGCGCCACTCG	600
QY	201	ProAlaLeuGlyArgGlnValValAspMetGlyIleAlaValGlnThrPheSerAlaLeu	220
Db	601	CCAAGCCTTGCTGACAGAGTTCTTGACACAGGGGGTTCGGTTCAAGCTACTCGGGCGC	660
QY	221	AsnValValArgThrValLeuAlaProAlaLeuAlaSerArgProSerValGlnGlyAla	240
Db	661	AACGCCGTACGTACCGTATTGGCTCCGGCACTGGCGCTCCAGACCCGCGTGACAGGTGCT	720
QY	241	ValAspPheGlyValSerThrAlaGlyGlyLeuValAlaAsnAlaGlyPheGlyAspArg	260
Db	721	GTTGACCTTGTGTATCGATGGCGGGTGTGCTGCTGCCAACGACGAGGCTTTGGCAACGC	780
QY	261	MetLeuSerValGlnSerArgAspGlnLeuArgGlyGlyAlaPheValLeuGlyMetLys	280
Db	781	CTGCTCAGTGTGCAGTCGCGGTGATCACCACGCGTGGCGGTGCATTAGTCTCGGTTGAAG	840
QY	281	AspLysGluProLysAlaAlaLeuSerGluGluThrAspTrpLeuAspAlaTyrLysAla	300
Db	841	GATAAAGAGCCCAAGGCTCAACTGAGCGAAGAAACGACTGCTCGAGGCTTATAAGCA	900
QY	301	IleLysSerAlaSerTyrSerGlyAlaAlaLeuAsnAlaGlyLysArgMetAlaGlyLeu	320
Db	901	ATCAAAATCGGCCAGCTACTCGGGTGGCGGCTCAACGCTGGCAAGCGATGGCCGCTTG	960
QY	321	ProLeuAspValAlaThrAspGlyLeuLysAlaValArgSerLeuValSerAlaThrSer	340
Db	961	CCACTGGATATGGCGACCGACGCAATGGGTGGGTGAAGACCTGTGTACAGCGTCCAGC	1020
QY	341	LeuThrLysAsnGlyLeuAlaLeuAlaGlyLysTyrAlaGlyValSerLysLeuGlnLys	360
Db	1021	CTGACCCCAAAACGGTCTGCCCTTGCGCGGTGCTTGCACAGGGGTAGGCAAGTTGCAGAG	1080
QY	361	MetaIaThrLysAsnIleThrAspSerAlaThrLysAlaAlaValSerGlnLeuSerAsn	380

```

Db      1081  ATGGCGACGAAAAATATATCATCCGACCCGGCGACCAAGCGCGGTCAGTCAAGTGGACCAAC 1140
QY      381  LeuValGlySerValGlyValPheAlaGlyTrpThrThrAlaGlyLeuAlaThrAspPro 400
Db      1141  CTGGCAGGTTCCGGCAGCGCGTTTTCGCAGAGCTGGACCAAGCGCGCGCTGACAAACCGATCCC 1200
QY      401  AlaValLysLysAlaGluSerPheIleGlnAspLysValLysSerThrAlaSerSerThr 420
Db      1201  GCGGTGAAAAAGCCGAGTCTTCATACAGACACAGGTGAATCGACTGCATCCAGTACC 1260
QY      421  ThrSerTyrrValAlaAspGlnThrValLysLeuAlaLysThrValLysAspMetSerGly 440
Db      1261  ACAGGCTACGTAGCCGACCAAGCCGTCAAACTGGCGAAGACCGTCAAGACATGGCGCGG 1320
QY      441  GluAlaIleSerSerThrGlyAlaSerLeuArgSerThrValAsnAsnLeuArgHisArg 460
Db      1321  GAGGCGATCACCCATACCGCGCCCAAGCTTGGCAATACGGTCAATAAACCCTGCGTCAACGC 1380
QY      461  SerAlaProGluAlaAspIleGluGlyGlyIleSerAlaPheSerArgSerGluThr 480
Db      1381  CCGGCTCGTGAAGCTGATATAGAAGAGGGGGCAGCGCGCT--TCTCCAAGTGAATA 1437
QY      481  ProPheGlnLeuArgArg 486
Db      1438  CCGTTTCGGCCTATGCGG 1455

RESULT 3
US-09-825-414-1/c
; Sequence 1, Application US/09825414
; Patent No. US20020083489A1
; GENERAL INFORMATION:
; APPLICANT: Colimer, Alan
; APPLICANT: Alfano, James R.
; APPLICANT: Charkowski, Amy O.
; TITLE OF INVENTION: DNA MOLECULES AND POLYPEPTIDES OF PSEUDOMONAS SYRINGAE
; FILE REFERENCE: 19603/3243
; CURRENT APPLICATION NUMBER: US/09/825,414
; CURRENT FILING DATE: 2001-04-03
; PRIOR APPLICATION NUMBER: 60/194,160
; PRIOR FILING DATE: 2000-04-03
; PRIOR APPLICATION NUMBER: 60/224,604
; PRIOR FILING DATE: 2000-08-11
; PRIOR APPLICATION NUMBER: 60/249,548
; PRIOR FILING DATE: 2000-11-17
; NUMBER OF SEQ ID NOS: 91
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 30365
; TYPE: DNA
; ORGANISM: Pseudomonas syringae
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (29734)
; OTHER INFORMATION: n at any position is undefined
US-09-825-414-1

Alignment Scores:
Pred. No.: 1.95e-182 Length: 30365
Score: 1869.50 Matches: 381
Percent Similarity: 85.19% Conservative: 33
Best Local Similarity: 78.40% Mismatches: 71
Query Match: 77.67% Indels: 1
DB: 10 Gaps: 1

US-09-825-414-66 (1-487) x US-09-825-414-1 (1-30365)
QY      1  MethIStleasnGlnSerAlaGlnGlnProProGlyValAlaMetGluSerPheArgThr 20
Db      23984  ATGCACATCAACCGACGCGCTCCACAACACCGCGCTGTGACTGCGACGGATAGCTTTGGGACA 23925
QY      21  AlaseraSpAlaSerLeuAlaSerSerSerSerValArgSerValSerThrThrSerCysArg 40

```

D	b	23924	GCGTCCGACGGCTCTTGCCCTCCAGCTCTGTGCATCTGTCCAGCTCCGATCAGCAACGC				23865
O	y	41	AspleuglAlalIeThrAspTryLeuLysHisHisValPheAlaAlaHisArgPheSer				60
D	b	23864	GAGATAAATGCGATTGCCGATTACTGACAGATCATGTGTTCGCTGCCCATTAAC TGCCG				23805
O	y	61	VallIeGlySerProAspGluaArgAspAlaAlaLeuAlaHisasnGLuInIleAspAla				80
D	b	23804	CCGGCCGATTCCGGCTGATGGCCAAAGCTGCAGTTGACGTACACAATGGCCAGATCACTGGC				23745
O	y	81	LeuValGluThrArgrAlaAsnArgLeuTyrrSerGluGlyGluThrProAlatrIleAla				100
D	b	23744	CTGATTCGAGACGCGCGCCAGCCGCTGCACCTTCGAAGGGGAACC CGGCAACCATCCGC				23685
O	y	101	GluThrPheAlaLysAlaGluLysPheaspArgLeuAlaThrThrAlaSerSerAlaPhe				120
D	b	23684	GACACCTTCGCCCAAGCGGAAAAGCTTCACCCGATTGGCCAGCACATCATCAGCGCGTTG				23625
O	y	121	GluAsnThrProPheAlaAlaAlaSerValleuGlnTyrrMetGlnProAlaIleasnLys				140
D	b	23624	CGGGCGACGCCCTTTGCCATGGCCTGCTTGTCTTAGTACATGCAGCCCTGCGATCAACAAG				23565
O	y	141	GlyAspTrpLeuAlaThrProleuLysProleuThrProleuIleSerGlyAlaLeuSer				160
D	b	23564	GCGGATTGGCTGCCGCTCCGCTCAAACCGCTGACCCCGCTCATTTCCGGAGCGCTGTGC				23505
O	y	161	GlyAlaMetaspGlnValGlyThrLysMetMetaspArgAlaArgGlyAspLeuHisTyrr				180
D	b	23504	GCGGCCATGGACCAAGGTGGGCACCAAGATGATGACCGCGCAGCGGTGATTCGATTAC				23445
O	y	181	LeuSerThrSerProAspLysLeuHisaspAlaMetAlaValSerValLysArgHisSer				200
D	b	23444	CTGAGCGCCTCGCCGACAGGCTCCAAGATGCATGGCCGCTTCGGTGAAGCGCCACTCG				23385
O	y	201	ProAlaLeuGlyArgGlnValValaspMetGlyIleAlaValGlnThrPheSerAlaLeu				220
D	b	23384	CCAAGCCTTGCTCGACAGGTTCTGGACACGGGGGTTGCGGTTCAAGACGTACTGGCGCGC				23325
O	y	221	AsnValValArgThrValLeuAlaProAlaLeuAlaSerArgProSerValGlnGlyAla				240
D	b	23324	AACGCCGTACGTACCGTATTGGCTCGGCACTGGCGTCCAGACCCGCGCTGCAGGGTGCT				23265
O	y	241	ValaspPheGlyValSerThrAlaGlyGlyLeuValAlaasnAlaGlyPheGlyaspArg				260
D	b	23264	GTTGACCTTGGTGTATCGATGGCGGGTGTGCTGCCAACGACGAGCTTTGGCAACGC				23205
O	y	261	MetLeuSerValGlnSerArgaspGlnLeuArgGlyGlyAlaPheValIleuGlyMetLys				280
D	b	23204	CTGCTCAGTGTGCAGTCCGCGTATCAACCAAGCGTGGCGGTGCATTAGTGCCTCGGTTGAAG				23145
O	y	281	AspLysGluProLysAlaAlaLeuSerGluGluThrAspTrpLeuAspAlaTyrrLysAla				300
D	b	23144	GATAAAGAGCCCCAAGGCTCAACTGAGCGAAGAAAAACGACTGGCTCGAGGCTTATAAAGCA				23085
O	y	301	IleLysSerAlaSerTyrrSerGlyAlaAlaLeuAsnAlaGlyLysArgMetAlaGlyLeu				320
D	b	23084	ATCAAAATCGGCCAGCTACTCGGGTGGCGGCTCAACGCTGGCAAGCGGATGGCGGTTCTG				23025
O	y	321	ProLeuAspValAlaThrAspGlyLeuLysAlaValArgSerLeuValSerAlaThrSer				340
D	b	23024	CCACTGGATATGGCGACGACGCAATGGGTGCGGTAAGAAGCCTGGTGTCAAGGTTCCAGC				22965
O	y	341	LeuThrLysAsnGlyLeuAlaLeuAlaGlyGlyTyrrAlaGlyValSerLysLeuGlnLys				360
D	b	22964	CTGACCCAAAACGCTCGGCCCTGGCGGGTGCTTTCAGAGGGGTAGGCAAGTTCCACGAG				22905
O	y	361	MetaLathrLysAsnIleThrAspSerAlaThrLysAlaAlaValSerGlnLeuSerAsn				380
D	b	22904	ATGGCGACGAAAAATATACCCGACCCGGCGGCAAGGCCGGTCAGTACAGTTGACCAAC				22845
O	y	381	LeuValGlySerValGlyValPheAlaGlyTrpThrThrAlaGlyLeuAlaThrAspPro				400

Db	22844	CTGGCAGGTTCCGCACGCCGTTTTTCGCAGAGCGTGGACCACGCCGCCGTGACAACCAATCCC	227
QY	401	AlaValLysLysAlaGluSerPheIleGlnAspLysValLysSerThrAlaSerSerThr 	420
Db	22784	GCGGTGAATAAAAGCCGAGTGCTTCATACAGGACACCGGTGAATCGACTGCATCCAGTACC	2272
QY	421	ThrSerTyrrValAlaaspGlnThrValLysLeuAlaLysThrValLysAspMetSergly 	440
Db	22724	ACAGGCTACGTAGCCGACACAGACCGCTCAAACTGGCGAAGAACCCTCAAGACATGGCGGG	2266
QY	441	GluAlaIleSerSerThrGlyAlaSerLeuArgSerThrValAsnAsnLeuArgHisArg 	460
Db	22664	GAGCGCATCACCATACCGCGCCAGCTTGCGCAATACGGTCAATAACCTGCGTCAACGC	2260
QY	461	SerAlaProGluAlaAspIleGluGlyGlyIleSerAlaPheSerArgSergluThr 	480
Db	22604	CCGCGCTCGTGAAGCTGTATTAGAAAGAGGGGGGCACGGCGGCT--TCTCCAAGTGAATA	2254
QY	481	ProPheGlnLeuArgArg 486 	
Db	22547	CCGTTTCGGCCTATGCGG 22530	
<hr/>			
RESULT 4			
US-09-971-536-27			
; Sequence 27, Application US/09971536			
; Patent No. US20020159976A1			
; GENERAL INFORMATION:			
; APPLICANT: Glenn, Matthew			
; APPLICANT: Havukkala, Ilkka			
; APPLICANT: Bloksberg, Leonard			
; APPLICANT: Lubbers, Mark			
; APPLICANT: Dekker, James			
; APPLICANT: Christensson, Anna			
; APPLICANT: O'Toole, Paul			
; APPLICANT: Reid, Julian			
; APPLICANT: Coolbear, Timothy			
; TITLE OF INVENTION: Lactobacillus rhamnosus Polynucleotides, Polypeptide			
; FILE REFERENCE: 1043c2			
; CURRENT APPLICATION NUMBER: US/09/971,536			
; PRIOR FILING DATE: 2001-10-02			
; PRIOR APPLICATION NUMBER: U.S. No. US20020159976A1 09/634,238			
; PRIOR FILING DATE: 2000-08-08			
; PRIOR APPLICATION NUMBER: U.S. No. US20020159976A1 09/724,623			
; PRIOR FILING DATE: 2000-11-28			
; PRIOR APPLICATION NUMBER: PCT/NZ01/00160			
; PRIOR FILING DATE: 2001-08-08			
; NUMBER OF SEQ ID NOS: 83			
; SOFTWARE: FastSeq for Windows Version 4.0			
; SEQ ID NO 27			
; LENGTH: 4645			
; TYPE: DNA			
; ORGANISM: Lactobacillus rhamnosus			
US-09-971-536-27			
<hr/>			
Alignment Scores:			
Pred. No.: 2.16e-05 Length: 4645			
Score: 147.00 Matches: 104			
Percent Similarity: 33.53% Conservative: 64			
Best Local Similarity: 20.76% Mismatches: 241			
Query Match: 6.11% Indels: 92			
DB: 9 Gaps: 12			
<hr/>			
US-09-825-414-66 (1-487) x US-09-971-536-27 (1-4645)			
QY	1	MethisILeAnGInSeRaLaGlNGlNPrOProGIyVaLaAMetGLUSeRPhEArGTThr ::: ::: ::: :::	20
Db	2213	ATCAACGCCAATAGTGTCTGATGTCAACAGCTTCACAGGCAAAAGGCTACAGAGTGTCTGCT	2272
QY	21	AlaSerAspAlaSerLeuAlaSerSerSerValArgSerValSerThrThrserysArg 	40
Db	2273	GCTAGTGAATGCGCGAGTTATGCAAGTGAAGCGCAATCGATTGCTGGCAGTCAATGCTGAT	2332

QY 41 AspleuGlnAlaIleThrAspTyrLeuLysHisHisValPheAlaAlaHisArgPheSer 60
Db 2333 AATATGGAA-----ATCAAGTCTCTCGCCAGTGATGCTGAGAACAAATCG 2377
QY 61 ValIleGlySerProAspGluArgAspAlaAlaLeuAlaHisAsnGluGlnIleAspAla 80
Db 2378 CAAATTGCTTTGGCAGCTAGCAAGTCTGTGGCTAGTTCAGTGGCGAGCGCTCCGCA 2437
QY 81 LeuValGluThrArgAlaAsnArgLeuTyrSerGluGluGluThrProAlaThrIle--- 99
Db 2438 GCAATCGTG-----GCAAGTAGCGCGGCTAGTGAAGCGTCATCTGCAGCTGCTGCCGTA 2491
QY 100 -----AlaGluThrPheAlaLysAlaGluLysPheAspArgLeuAlaThrThr 115
Db 2492 AGTAACGCTGATGCATCAGCAAACTCTGCAGCGCGCTTATGATTCCTACGCTTCTGAG 2551
QY 116 AlaSerSerAla-----PheGluAsnThrProPheAlaAla 127
Db 2552 GCCAGTGCCGCTTCTGCTGCTAATGATAGTTCGGGATATGCCACTGCATCATTTGCAAGCA 2611
QY 128 AlaSerValLeuGlnTyrMetGlnProAlaIleAsnLysGlyAspTyrPheAlaThrPro 147
Db 2612 AGTTCGGCTGGCTGCCATG----- 2632
QY 148 LeuLysProLeuThrProLeuIleSerGlyAlaLeuSerGlyAlaMetAspGlnValGly 167
Db 2633 -----AGCGCAGCGCTTATCGACAGCGCAAGTTGCTGCCAAG 2668
QY 168 ThrLysMetMetAspArgAlaArgGlyAspLeuHisTyrLeuSerThrSerProAspLys 187
Db 2669 GTTGCAGTGATGATGCACAGCAGCGGGT----- 2698
QY 188 LeuHisAspAlaMetAlaValAlaSerValLysArgHisSerProAlaLeuGlyArgGlnVal 207
Db 2699 -----AGTGCAGCTGCTGTTGCTAGTGCAGCTCAAAAGCGACTCCAGAATTAACAAGCG 2752
QY 208 ValAspMetGlyIleAlaValAlaGlnThrPheSerAlaLeuAsnValValArg-----Thr 225
Db 2753 ACTGCAGCTACAGCAAGAAGTCAAGCACTTGATGATTGAAATAGATCAAGTCTCTAACT 2812
QY 226 ValLeuAlaProAlaLeuAlaSerArgProSerValGlnGly----- 239
Db 2813 GATTACGCAAGTGGCGCAAGCTCCAGTCCAGCGGAAGCGGCTCAAGCATGCACTGCCAACA 2872
QY 240 -----AlaValAspPheGlyValSerThrAlaGlyGlyLeuValAlaAsn 254
Db 2873 TCTGCGTATGCTAGTGTGCAAGTTGAGTGGCCAGTGAAGCCGGTTCATATGCTCATCAG 2932
QY 255 AlaGlyPheGlyAspArgMetLeuSerValGlnSerArgAspGlnLeuArgGlyGlyAla 274
Db 2933 GCAGGCTCCAGCGCCAGTAGCAGCTGTGCGTCACTCCGGCAGTAGCAGCCCAACATGCCAGC 2992
QY 275 PheValLeuGlyMetLysAspLysGluProLysAlaAlaLeuSerGluGluThrAspTyr 294
Db 2993 ACCGCTGCGAGTGGCGCATCCAGCTATCCGAGAGATAGTGGGATTCAGTCACTAACCCAGT 3052
QY 295 LeuAspAlaTyrLysAlaIleLysSerAlaSerTyrSerGlyAlaAlaLeuAsnAlaGly 314
Db 3053 CAGGCTGCAGCGCAGCAGCAAAAGGCAAGCAGTAACGCGAGTGCCGCAACCAACGCCGCC--- 3109
QY 315 LysArgMetAlaGlyLeuProLeuAspValAlaThrAspGlyLeuLysAlaValArgSer 334
Db 3110 -----GCGCGCGTGTGTTTCAGTGTCTGCCAGTGATGCAAGTGAACAGCGGAAGAGC 3160
QY 335 LeuValSerAlaThrSerLeuThrLysAsnGlyLeuAlaLeuAlaGlyIleTyrAlaGly 354
Db 3161 GCTGCAGTGCCTGATGTGTGCAAGCAGTGGCGCCAGCACGCTAACAGTAATGCGAGT 3220
QY 355 ValSerLysLeuGlnLysMetAlaThrLysAsnIleThrAspSerAlaThrLysAlaAla 374
Db 3221 GCCGCAGCC-----AGTGCAGACCAAGCGCTGGT 3247

QY 375 ValSerGlnLeuSerAsnLeuValGlySerValGlyValPheAlaGlyTyrPThrThrAla 394
Db 3248 GATAGCAAAAGCCGCA-----GCAGATTCCTCGAGTGA 3280
QY 395 GlyLeuAlaThrAspProAlaValLysLysAlaGluSerPheIleGlnAspLysValLys 414
Db 3281 GCGAGTGCTGCAGCAGCAGTAGTCCCAAGGGTGCAGACAGTAGTCCAGCGAAGCGCGAGT 3340
QY 415 SerThrAlaSer-----SerThrThrSerTyrValAlaAspGlnThrValLysLeuAla 432
Db 3341 GCCGCGGCATCCGATGACTCGGTAGCTTCTAGTCCCGCCAGTGCGGCTGCAGGCTTTGAC 3400
QY 433 LysThrValLysAspMetSerGlyGluAlaIleSerSerThrGlyAlaSerLeuArgSer 452
Db 3401 AAAGCTGCAGCGCTCGGGAAGCGCGCAGCTTCAAGTGC CGGAGCGGCTGCTAGTTCA 3460
QY 453 ThrValAsnLeuArgHisArgSerAlaProGluAlaAspIleGluGluGlyIle 472
Db 3461 GCGCGAGCTCAAGCAGCAGCAGAGTGGCGCAAGCTCCAGTCCAGCGAAGCGGCTCAAGCA 3520
QY 473 Ser 473
Db 3521 TCA 3523

RESULT 5
US-09-813-214A-8
; Sequence 8, Application US/09813214A
; Patent No. US20020177200A1
; GENERAL INFORMATION:
; APPLICANT: Tucker, Kenneth
; APPLICANT: Plosila, Laura
; TITLE OF INVENTION: MORAXELLA CATARRHALIS OUTER MEMBRANE PROTEIN-106 POLYPEPTIDE,
; FILE REFERENCE: 7969-089-999
; CURRENT APPLICATION NUMBER: US/09/813, 214A
; PRIOR APPLICATION NUMBER: 2000-03-20
; PRIOR FILING DATE: 1997-11-12
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 8
; LENGTH: 9542
; TYPE: DNA
; ORGANISM: Moraxella catarrhalis
US-09-813-214A-8

Alignment Scores:
Pred. No.: 6 75e-05 Length: 9542
Score: 146.50 Matches: 111
Percent Similarity: 36.00% Conservative: 69
Best Local Similarity: 22.20% Mismatches: 181
Query Match: 6.09% Indels: 139
DB: 9 Gaps: 23

US-09-825-414-66 (1-487) x US-09-813-214A-8 (1-9542)
QY 32 ArgSerValSerThrThrSerCysArgAspLeuGlnAlaIleThrAspTyrLeuLysHis 51
Db 1346 AAGACTTAACTATCAACAGGTGGTGCACAGACAGCAGCGCATTAACCGAT----- 1393
QY 52 HisValPheAlaAlaHisArgPheSerValIleGlySerProAspGluArgAspAlaAla 71
Db 1394 -----CATACATCGGTGTGTACAAATGGCGATGCTGTGAAGTTCAA 1438
QY 72 LeuAlaHisAsnGluGlnIleAspAlaLeuValGluThrArg-----AlaAsnArg 88
Db 1439 CTGCTGAAACTTTAACCAAGCCTTAAATGGTTTACCACTGAATAACCGCAACGAG 1498
QY 89 LeuTyrSerGluGlyGluThrProAlaThrIleAlaGluThrPheAlaLysAlaGluLys 108
Db 1499 AAAGTTACCGTAGGCAAAAC----- 1519
QY 109 PheAspArgLeuAlaThrThrAlaSerSerAlaPheGluAsnThrProPheAlaAlaAla 128

Db 1520 -----CGCCTT---ACCACAGATAAAATTGGTTTACCAATGAT-----ATG 1558
OY 129 SerValLeuGlnTyrMetGlnProAlaIleAsnLysGlyAspTyrPheAlaThrProLeu 148
Db 1559 AATGGCATGTGATGAAGCAACACCTTATCTTGATAAAGACACTGGCATTCATGCAGGTGCT 1618
OY 149 LysProLeuThrProLeuIleSerGlyAlaLeu----- 159
Db 1619 CAAAAGATTACCAAACTTACTGCTGGTGTAGTAGATGACGATGGCGCAACTTATGCACAG 1678
OY 160 -----SergIlyAlaMetAspGlnValGlyThrLysMetMet 171
Db 1679 CTTAAAAAGTTAAACCAACCGCTGAAGTGTCTACAAACCTTTACCGCTTAAAAAGGTA 1738
OY 172 AspArgAlaArgGlyAspLeuHisTyrLeuSerThrSerProAspLysLeuHisAspAla 191
Db 1739 GATAAAAATGCTAATGAT-----GCTAATGACAGC 1768
OY 192 MetAlaValSerValLysArgHisSerProAlaLeuGlyArgGlnValValAspMet--- 210
Db 1769 AAAATCATCATCCGTGGGTAATAAATAACAACACGAGCGGTACTCATGCAACACCCCTAAAA 1828
OY 211 -----GlyIleAlaValGlnThrPheSerAlaLeuAsnValValArgThr 225
Db 1829 CTCAAAGGTGAACCGGTGTGATGTTACAACC----- 1861
OY 226 ValLeuAlaProAlaLeuAlaSerArgProSerValGlnGlyAlaValAspPheGlyVal 245
Db 1862 -----GAAACAATGTGTACAGTTACCTTTGGGCTT 1891
OY 246 SerThrAlaGlyGlyLeu---ValAlaAsnAlaGlyPheGlyAspArgMetLeuSerVal 264
Db 1892 AACCAAAATAACGGCTCTGACCGTTGGCAACAGCACCCCTAAACAACGATGGCTTATCTGTT 1951
OY 265 Gln-----SerArgAspGlnLeuArgGlyGlyAlaPheValLeuGlyMetLysAsp 281
Db 1952 AAAAACACCAATAGTAACAACAATCCAAATCGGTGCTGATGGCATTTACTTACTGAT 2011
OY 282 -----LysGluProLysAlaAlaLeuSerGluGluThr-----AspTyr 294
Db 2012 ATCAGCAATAGTAAGCCAGGTGCTGGCATTGAAATATACCACCTCCGATTACACGAGACGGT 2071
OY 295 LeuAspAlaTyrLysAlaIleLysSerAlaSerTyrSerGlyAlaAlaLeuAsnAlaGly 314
Db 2072 ATT-----GGTTTGTCTAATATACTGCT 2095
OY 315 LysArgMetAlaGlyLeuProLeuAspValAlaThrAsp-----GlyLeuLysAla 331
Db 2096 TCATTTGATGCAAAACAACCCCGCTTAACCCCAACTGGCATTTAACGCGAGGTGCTAAAGAG 2155
OY 332 ValArgSerLeuValSerAlaThrSerLeuThrLysAsnGlyLeuAlaLeuAlaGlyGly 351
Db 2156 CTGACCAATGTCCAATCTGCCATTAACCTGCTACCAATGCT----- 2197
OY 352 TyrAlaGlyValSerLysLeuGlnLysMetAlaThrLysAsnIleThrAspSerAlaThr 371
Db 2198 ---GGGCAAGTAGACTTTATGACACCGCCTAAGCACTGCTAATATACGAAAAATCAGGCTCT 2254
OY 372 LysAlaAlaValSerGlnLeuSerAsnLeuValGlySerValGlyValPheAlaGlyTyr 391
Db 2255 GCCGCCACCATTTAAAGACTTATACACCTATTCACAAGTACCGCTGACCTTTGCAGGT--- 2311
OY 392 ThrThrAlaGlyLeuAlaThrAspProAlaVal---LysLysAlaGluSerPheIleGln 410
Db 2312 -----GATACAGGTCTTAATGTCAACCAAAAACTGGCGGAGATTTTAAAG 2356
OY 411 AspLys---ValLysSerThrAlaSerSerThrThr-----SerTyrValAla 425
Db 2357 GTTAAAGGTGCTAAACACACAGCTGATGATTTAACCAAAAAATAACATCGGTGTGTGCT 2416
OY 426 AspGln-----ThrValLysLeuAlaLysThrValLysAspMetSergly 440

Db 2417 GATAGTACCGATAATAGCTTAACCGTTAAACTTGCTTAAACCTTTAAGCGATCTGATGCC 2476
OY 441 GluAlaIleSerSerThrGlyAlaSerLeuArgSerThrValAsnAsnLeuArgHisArg 460
Db 2477 GTTAATACTAAACCCCTACTGTCGACGATAAAGTTACCGTAGACGTGCACAAC--- 2533
OY 461 SerAlaProGluAlaAspIleGluGluGlyGlyIleSerAlaPheSerArgSergluThr 480
Db 2534 -----ACCGCTAAGCTACAAAAATGCTGATTTAACC---TTTAGCAAAACAATAFACA 2581
RESULT 6
US-09-932-183A-1
; Sequence 1, Application US/09932183A
; Patent No. US20020127641A1
; GENERAL INFORMATION:
; APPLICANT: Estell, David A.
; TITLE OF INVENTION: Proteases From Gram-Positive Organisms
; FILE REFERENCE: GC394C1-US
; CURRENT APPLICATION NUMBER: US/09/932,183A
; PRIORITY FILING DATE: 2001-08-17
; PRIORITY APPLICATION NUMBER: US 09/308,375
; PRIORITY FILING DATE: 1999-05-14
; PRIORITY APPLICATION NUMBER: PCT/US98/18828
; PRIORITY FILING DATE: 1998-09-08
; PRIORITY APPLICATION NUMBER: EP9719636.4
; NUMBER OF SEQ ID NOS: 3
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 7100
; TYPE: DNA
; ORGANISM: Bacillus subtilis
US-09-932-183A-1
Alignment Scores:
Pred. No.: 0.000209 Length: 7100
Score: 140.00 Matches: 128
Percent Similarity: 33.44% Conservative: 81
Best Local Similarity: 20.48% Mismatches: 222
Query Match: 5.82% Indels: 194
DB: 10 Gaps: 26
US-09-825-414-66 (1-487) x US-09-932-183A-1 (1-7100)
OY 15 MetGluSerPheArgThrAlaSerAspAlaSerLeuAlaSerSerSerValArgSerVal 34
Db 862 ATGGAATTATATCAGCGACAGGACAGCAATAATGTTCAAAACCTTAATAATACACGGTATGCC 921
OY 35 SerThrThrSerCysArgAspLeuGlnAlaIleThrAspTyrLeuLysHisHisValPhe 54
Db 922 AGTTCTATGGGCTCTAGTAATAGACAGCTGTTCAAGATTATTTGAAT----- 969
OY 55 AlaAlaHisArgPheSerVal---IleGlySerProAspGluArgAspAlaAlaLeuAla 73
Db 970 GCAGTAATAGTCTTATGTAAAGCACTGGAAGCAATATATATGACATCACAATTCAAAGC 1029
OY 74 HisAsnGluGlnIleAspAlaLeu-----ValGluThrArgAlaAsnArgLeuTyr 90
Db 1030 TTGAATATGCAATTTAGAGAATTAGCCTCCAAACGCTCAACACAGCTGCTAATCAAGCTCT 1089
OY 91 SerGluGlyGluThrProAlaThrIleAlaGluThrPhe----- 103
Db 1090 TCTTTTGA-----GCAGAACTAACCCCAACCTTCAAAAGCATGTCACCTATTTTA 1140
OY 104 -----AlaLysAla 106
Db 1141 ATCTCCGTTCTTATTCTACGAGCTATCTGACTTAAGAAATGCTATCCACGCA 1200
OY 107 GluLysPheAspArgLeuAlaThrThrAlaSerSerAlaPheGluAsnThrProPheAla 126
Db 1201 ATAGAAATTGATCTCTCATGACAAATAFTGCCGCTGTATGAATGAGCCGATTTATAA 1260
OY 127 AlaAlaSerValLeuGlnTyrMetGlnProAlaIleAsnLysGlyAspTyrPheAlaThr 146

Db 1261 TATAATGAACCTCTC-----CAAGAACTCTATTGACTTAGGTGATACACTTTCAAAT 1311
QY 147 PROLEULysPROleuThrPROleuIleSerGlyAlaIleuSerGlyAlaIleuSerGlnVal 166
Db 1312 AAATCACAGATATTCTTCAAAATGACAGCGCATTTTGGAGAATGGGTTTCAGTGAAGT 1371
QY 167 GlyThrLysMetMetAspArgAlaArgGlyAspLeuHisTyrLeuSer---ThrSerPro 185
Db 1372 GAGCTCTCCACGTTAAGCAAACTGCCCAAGTTCTTCAAAATGTCTCTGATTTAACCTCC 1431
QY 186 AspLysLeuHisAspAlaMetAlaValSerValLysArgHisSerProAlaIleuGlyArg 205
Db 1432 GATGATACAGTTAAACACTTAACGGCAGCAATGCTCAACTTTTAATATATGCGACCAATGAT 1491
QY 206 GlnVal-----ValAspMetGlyIleAlaValGlnThrPhe 217
Db 1492 TCAATATCAATTCAGATTAATTAATGAGGTTGATATAACTATGCTGTACAACTCTA 1551
QY 218 SerAlaIleuAsnValValArgThrValIleuAlaProAlaIleuAlaSerArgProSerVal 237
Db 1552 GATCTGGCCAATTCTATCGCTAAA-----GCTGGTTCAACTGCTTCTACATTGCGGGTA 1605
QY 238 GlnGlyAlaValaAspPheGlyValSerThrAla-----Gly 249
Db 1606 GAGCTAAATGATCTTATTGTTGTTATACAACTGCAATTGCTAGTACACACAGTGAATCAGGG 1665
QY 250 GlyLeuValAlaAsn-----AlaGlyPheGlyAspArgMetLeuSer 263
Db 1666 AATATCGTCGGAACTCCTTAAGACAAATTTTCGCGCGATTTGGGAATTAATCAAAAGCTCA 1725
QY 264 ValGlnSerArgAspGlnLeuArgGlyGlyAlaPheValIleuGlyMetLysAspLysGln 283
Db 1726 ATTAAGCGTTAGAACAGATGGTATCTCAGTTAAACAGCTGGTGTGAAGCTAAATCA 1785
QY 284 ProLysAlaAlaLeuSerGln-----GluThrAspTrpLeu---AspAlaTyrLys 299
Db 1786 GCAAGTGATTTAATTAGTGAAGTGTGCTGTAAGTGGGATACGCTTTCGATGCTCAGAAA 1845
QY 300 AlaIleLysSerAlaSerTyrSerGly----- 308
Db 1846 CAAAATACTCTCAATTGGAGTAGCTGTATTATCAATTATCCCGTTTAAATGCAATGATG 1905
QY 309 -----AlaAlaIleuAsnAlaGlyLys----- 315
Db 1906 AACCACTTCTCTATTGCTCAGAATGCGGCTAAAACTGCGGCTAACTCAACAGAGAAGTGCT 1965
QY 315 ----- 315
Db 1966 TGGAGTGAGCAGCAAAAGTATGACAGTAGTCTACAAAGCTAAGGTTAAATGAAGCTCAAAAT 2025
QY 315 ----- 315
Db 2026 AACTTCACTGAATTGCTATTGACAGCTTCTGATGCTTTATTAGCGAGCGGATTAAATTGAA 2085
QY 316 -----ArgMetAlaGlyLeuProLeuAspValAlaThrAspGlyLeuLysAlaVal--- 332
Db 2086 TTTACTCAAGCCGACGGTCTTTGCTTAACGCTTCTACAGAGAGTAATCAAAATCAGTTGGG 2145
QY 333 -----ArgSerLeuValSerAlaThrSer-----LeuThrLysAsnGly 345
Db 2146 TTCTTACCTCCCTTTTACTGTCAGTAAGCAGCTGCAACCTTTTGTCTCAGTAAGATTAACC 2205
QY 346 LeuAlaIleuAlaGly-----GlyTyrAlaGlyValSerLysLeuGlnLysMet 361
Db 2206 CGCACATTTAGCCAGCAGCCTAATTTTGGGCACACAGCTGCAATGGGGCAA---GAACCTTTTA 2262
QY 362 AlaThrLysAsnIleThrAspSerAlaThrLysAlaAlaValSerGln----- 377
Db 2263 GCGACTGCTGGGCTAGAAAGCTGGTATGACTCGTGACAGAGTCCGCCCTCAAGAGTTCTTAAA 2322
QY 378 -----LeuSerAsnLeuValGlySerValGlyValPheAlaGly 390

Db 2323 ACTGCTCTTCGAGGGTTGCTTGTTCACACTTTAGTTGGCGGTGCATTTGCTGCTTTGGGA 2382
QY 391 TrpThrThrAlaGlyLeuAlaThrAspProAla---ValLysLysAla----- 405
Db 2383 TGGGCGCTAGAAATCATTAATTCTTCTTCTTTGACAGAGCTAAAAAGCTAAAGATGATTTT 2442
QY 406 -----GluSer 407
Db 2443 GAGCAGAGCCAGCAAAACCAATGTCGAAGCAATTACGACCAATAAAGACTCCACTGATAAA 2502
QY 408 PheIleGln-----AspLysValLysSerThrAlaSerSerThrThrSer 422
Db 2503 CTAATACAGCAATATTAAGAGCTTCAAAAAGTTAAAGAGTCAAGATCTTTAACTTCAGAT 2562
QY 423 TyrValAlaAspGln-----ThrValLysLeuAlaLysThr-----Val 435
Db 2563 GAAGAGCAAGAATACCTTCAAGTCACCTCAGCAATTAGCACAACCTTCCCTGCAATTAGTT 2622
QY 436 Lys-----AspMetSerGlyGluAlaIleSerSerThrGlyAlaSerLeuArgSerThr 453
Db 2623 AAAGCTATGATTCTCAAGGAATGCAATTCTTAAGACAAATAAAGAGCTTGAANAAGCG 2682
QY 454 ValAsnAsnLeuArg 458
Db 2683 ATTGAGAATTAATAA 2697

RESULT 7

US-09-738-626-2739
; Sequence 2739, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738, 626
; PRIOR FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 2739
; LENGTH: 8979
; TYPE: DNA
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-2739

Alignment Scores:
Pred. No.: 0.000469 Length: 8979
Score: 138.00 Matches: 119
Percent Similarity: 37.16% Conservative: 85
Best Local Similarity: 21.68% Mismatches: 219
Query Match: 5.73% Indels: 126
DB: 9 Gaps: 23

US-09-825-414-66 (1-487) x US-09-738-626-2739 (1-8979)

QY 7 AlacGlnGlnProGcGlyValAlaIleuMetGluSerPheArgThrAlaSerAspAlaSerLeu 26
Db 3073 GCACACACCGCACCTGACGTGCTGTGGCGCGCATGCGCCAGCAGATTTCGCTGCCGTA 3132

QY 27 AlaSerSerSerValArgSerValSerThrThrSerCysArgAspLeuGlnAlaIleThr 46
Db 3133 AAGTCCGCAGTCATCCACGACCGGATTCGGCATCCGTTGTGGAAGGCATGCTGCCCTG 3192
QY 47 AspTyrLeuYshHisValPheAlaAlaHisArgPheSerValIleGlySerProasp 66
Db 3193 GTTCACCTGGAGCACACAT-----GTGCTCAAGTCCGATGTC 3231
QY 67 GluArgAspAlaAlaLeuAlaHisAsnGlnGlnIleAspAlaLeuValGluThrArgAla 86
Db 3232 CCAACCGACGGCGCGTGAAGTTTCCGCGACTGCCGATGAGGTAGTCGATACCGACCTG 3291
QY 87 AsnArgLeuTyr-----SerGluGlyGluThrProAlaThr 98
Db 3292 GGTCCGCTCGTGCATCGTCCGCGCAGAAATCGCCGACGACGAGAAGCAACCTGATGCTACG 3351
QY 99 IleAlaGluThrPheAla---LysAlaGluLysPheAspArgLeuAlaThrThrAlaSer 117
Db 3352 TTGGCTGAGCGTTTCCGATCCGCGGACGCAAGGGCAACGCTGTCGACGCAACACACC 3411
QY 118 SerAlaPheGlu-----AsnThrProPheAlaAlaAlaSerValLeuGlnTyrMet 134
Db 3412 TCCGACATGCCAACCAACCGTGACACCCACGCTCAGCTCGCGCAGTGCACACCGTTGTT 3471
QY 135 GlnPro-----AlaIleAsnLysGlyAspTyrPleuAlaThrProleu 148
Db 3472 GCACCTGAATCCATGCCCATTCGCTGTGATCTCCGGTGAC----- 3513
QY 149 LysProLeuThrProleu-----IleSerGlyAlaLeuSerGlyAlaMetasp 164
Db 3514 -----CGCAACCCAAATTCACGTCTGTGATGTGCGGCTTCCCTGGCTGTCGCCAGT 3567
QY 165 GlnValGlyThrLysMetMetAspArgAlaArgGlyAspLeuHisTyrLeuSerThrSer 184
Db 3568 GTGATCGTGCACGGCATGTGACCTCTGCCATCGGTGACTGATCGCCGGTGCACGATTC 3627
QY 185 ProAspLysLeuHisAspAlaMetAlaValSerValLysArgHis----- 199
Db 3628 AACGATGAGCAGATCCAAACTCCCGCAGCCCAAGCTCGTGAATATACCCGCAACCATGCTG 3687
QY 200 SerProAlaLeuGlyArgGlnValValAspMetGlyIle-----AlaValGlnThr 216
Db 3688 GCACCAAGTCTTCCAGGTGAAGAAATGACTTACGCGTTGAGCGCTCCGACGTGACAAAC 3747
QY 217 PheSerAlaLeuAsnValValArgThrValLeuAlaProAlaLeuAlaSerArgProSer 236
Db 3748 CGCCACGAGATGGAGAGGTCCGACCGTTACCGCA-----ACC 3786
QY 237 ValGlnGlyAlaValAspPheGlyValSerThrAlaGlyGlyLeuValAla-----Asn 254
Db 3787 GTCAACGGCACTTA-----GTGCTTACCGCCACCGCTGTGTGGCAGCTCCATCT 3837
QY 255 AlaGlyPheGlyAspArgMetLeuSerValGlnSerArgAspGlnLeuArgGlyGlyAla 274
Db 3838 ACTTCTACGCATTCACAGGCCAGGCAATTCAGTCCACAGGC----- 3879
QY 275 PheValLeuGlyMetLysAspLysGluProLysAlaAlaLeuSerGluGluThrAspTyr 294
Db 3880 -----ATGGGTATGGAAGCAGCGCGTAATCTCAGGCAAGCTCGCGTATCTGGACCCG 3933
QY 295 LeuAspAlaTyrLysAlaIleLysSerAlaSerTyrSerGlyAlaAlaLeuAsnAlaGly 314
Db 3934 GCCGATGCACACACCCGCAATAAG-----CTGGGCTTCTCCATCGTG 3975
QY 315 LysArgMetAlaGlyLeuProLeuAspValAlaThrAspGlyLeuLysAlaValArg--- 333
Db 3976 GAAATCGTGAAAAACAACCCACGCGAATACCGTGGCAGGGAGAGATTCTTCCACCCA 4035
QY 334 ---SerLeuValSerAlaThrSerLeuThrLysAsnGlyLeuAlaLeuAlaGlyGlyTyr 352
Db 4036 GACGGCGTTTGTACCTCACCCAGTTCACCCAGGTGGCATGGCAACTCTG---GGCGTT 4092
QY 353 AlaGlyValSerLysLeuGlnLysMetAlaThrLysAsnIleThrAspSerAlaThrLys 372

Db 4093 GCTCAGATCGCTGAATAATGCGTGAAGCACATGCCCTTGAACCAAGCGTGCACTTGTCTGA 4152
QY 373 AlaAlaValSerGlnLeuSerAsnLeuValGlySerValGlyValPheAlaGlyTyrThr 392
Db 4153 CACTCCGTTGGTAGTACAACGCGCTTGCATATGCTGTGTGTGCTGTCCCTGGAATCC 4212
QY 393 Thr-----AlaGlyLeuAlaThrAspProAlaValLysLysAlaGlu 406
Db 4213 GTTCTGAGATCGTTTACCGCTCGTGGCTTGACCATGCACCGCTTGTGTGATCGGATGAA 4272
QY 407 SerPheIleGln-----AspLysValLysSerThrAla 417
Db 4273 AAGGCTCTGTCCAACCTACGCGCTCGCAGCTCTTCCGCCCAACAAGATGGGTCTGACCGCA 4332
QY 418 SerSerThrThrSerTyrValAlaAspGlnThrValLysLeuAlaLysThrValLysasp 437
Db 4333 GACAACGTTTTCGATTACGTTGCG-----TCTGTTCCGAA 4368
QY 438 MetSerGlyGlu-----AlaIle 443
Db 4369 GCTTCCGCTGAATTCCTGAGATCGTTAACTACACTTGCGCTGGCGCTGCAGTACGAGTT 4428
QY 444 SerSerThrGlyAlaSerLeuArgSerThrValAsnAsnLeuArgHisArgSerAlaPro 463
Db 4429 GCTGAACCCAGCGCTGCTT-----GCCGCCCTT 4458
QY 464 GluAlaAspIleGluGlu-----GlyGlyIleSerAlaPhe-----Serargser 478
Db 4459 CGTGCCGATGTGAGAACCGTGACACCAAGTCAAGCGTGCCTTCAATTGATCCCTGGCATT 4518
QY 479 GluThrProPheGlnLeuArgArgLeu 487
Db 4519 GACGTGCATTCACCTCCTCCAAAGCTG 4545

RESULT 8
US-09-379-931-6
; Sequence 6, Application US/09379931
; Patent No. US20020009792A1
; GENERAL INFORMATION:
; APPLICANT: Smit, John
; APPLICANT: Bingle, Wade H.
; APPLICANT: NO. US20020009792A1e1lin1, John F.
; TITLE OF INVENTION: EXPRESSION AND SECRETION OF HETEROLOGOUS POLYPEPTIDES FROM CA
; FILE REFERENCE: 08106/002003
; CURRENT APPLICATION NUMBER: US/09/379, 931
; CURRENT FILING DATE: 1999-08-24
; PRIOR APPLICATION NUMBER: US 08/614, 377
; PRIOR FILING DATE: 1996-03-12
; PRIOR APPLICATION NUMBER: US 08/194, 290
; PRIOR FILING DATE: 1994-02-09
; PRIOR APPLICATION NUMBER: US 07/895, 367
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 3300
; TYPE: DNA
; ORGANISM: Caulobacter crescentus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (101)...(3178)
US-09-379-931-6

Alignment Scores:
Pred. No.: 0.000475 Length: 3300
Score: 132.00 Matches: 101
Percent Similarity: 36.51% Conservative: 75
Best local Similarity: 20.95% Mismatches: 166
Query Match: 5.48% Indels: 140
DB: 10 Gaps: 23

US-09-825-414-66 (1-487) x US-09-379-931-6 (1-3300)

QY 22 SerAspAlaSerLeuAlaSerSerSerValArgSerValSerThrThrSerCysArgAsp 41
Db 224 TCGGACGCGCGCTGCGCTGACCAACACCCCTGAAGCTGGTCAACAGCAGCAGCGGCTGTGCC 283
QY 42 LeuGlnAlaIleThrAspTyrLeuLysHisHisValPheAlaAlaHisArgPheSerVal 61
Db 284 ATCCAGACC-----TACCAGTCTCTTCACCGCGCTGGCCCCGTCGGGCC 325
QY 62 IleGlySerProAspGluArgAspAlaAlaLeuAlaHisAsnGluGlnIleAspAlaLeu 81
Db 326 GCTGGT-----CTGGACTTCCCTG 343
QY 82 ValGluThrArgAlaAsn-----ArgLeuTyrSerGluGluGluThr 95
Db 344 GTCGACTCGACCACCAACACCAACGACCTGAACGCGCTACTACTCGAAG----- 394
QY 96 ProAlaThrIleAlaGluThrPheAlaLysAlaGluLysPhe-----AspArg 111
Db 395 -----TTCGCTCAGGAACCGCTTCATCATCACTTCTCGATCAAC 433
QY 112 LeuAlaThrThrAlaSerSerAlaPheGluAsnThrProPheAlaAlaAla----- 128
Db 434 CTGGCCACGGCGCGCGCGCGCGC-----GCGAGCGGCTTTCGCGCGCGCTACACGGGC 487
QY 129 ---SerValLeuGlnTyrMetGlnProAlaIleAsnLys-----GlyAspTrpLeuAla 145
Db 488 GTTTCGTACGCCAGACGGTGCACCGCCTATGACAAGATCATCGCAACGCCGCTCGCG 547
QY 146 ThrProLeuLysProLeuThrProLeuIleSerGlyAlaLeuSerGlyAlaMetAspGln 165
Db 548 ACC-----GCCGCTGGCGTGCAGCGTC 568
QY 166 ValGlyThrLysMetMetAspArgAlaArgGlyAspLeuHisTyrLeuSerThrSerPro 185
Db 569 GCGGCCGCGGCTTTCCTGAGCCGCCAGGCCAACATCGACTACCTGACC----- 619
QY 186 AspLysLeuHisAspAlaMetAlaValSerValLysArgHisSerProAlaLeuGlyArg 205
Db 620 -----GCCCTCGTGGCGGCCAACACGCGCTTCACGGCGCGCT 655
QY 206 GlnValAlaSpMetGlyIleAlaValGlnThrPheSerAlaLeuAsnValAlaArgThr 225
Db 656 GCCGACATCGATCTGGCGCTCAAGGCGCGCTGATCGGCACCATCTCGAACGCCGCCACG 715
QY 226 ValLeu-----AlaProAlaLeuAlaSerArgProSerValGlnGlyAlaValAsp 242
Db 716 GTGTCGGCATCGGTGGTTACGCGACCGCGCGCGATGATCAACGACCTGTCCGAC 775
QY 243 PheGlyValSerThr-----AlaGlyGlyLeu-----ValAlaAsnAla 255
Db 776 GCGGCCCTGTCGACCGACACGCGCGGTGGCGTGAACCTGTTCACCCGCCCTATCCGTCGTG 835
QY 256 GlyPheGlyAspArgMetLeuSerValGlnSer--ArgAspGlnLeuArgGly----- 272
Db 836 GCGGTGCGGGTTCGACCTCTCGTACCAACCGGCACCGACACCTGACGGGCACCGCC 895
QY 273 -----GlyAlaPheValLeuGlyMetLysAsp 281
Db 896 AACAAACGACACGTTGCTGCGGGTGAAGTCGCGCGCTGCGACCCCTGACCGTTGGCGAC 955
QY 282 LysGluProLysAlaAlaLeuSerGluGluThrAspTrpLeuAspAlaTyrLysAlaIle 301
Db 956 ACCCTGAGCGCGGTGCTGGCACCGACGCTCTGAACCTGGGTGCAAGCT----- 1003
QY 302 LysSerAlaSerTyrSerGlyAlaAlaLeuAsnAlaGlyLysArgMetAlaGlyLeuPro 321
Db 1004 -----GCTGCGGTACGCGT-----CTGCCG 1024
QY 322 LeuAspValAlaThrAspGlyLeuLysAlaValArgSerLeuValSerAlaThrSerLeu 341
Db 1025 ACCGGCGTGACGATCTCGGGCATCGAAACGATG---AACGTGACGTCGGCGCGCTGCGATC 1081

QY 342 ThrLysAsnGlyLeuAlaLeuAlaGlyGlyTyrAlaGlyValSerLysLeu----- 358
Db 1082 ACCCTGAAC-----ACGTCTTCGGCGGTGACGGGTCTGACCGCCCTGAACACCAAC 1132
QY 359 -----GlnLysMetAlaThrLysAsn 365
Db 1133 ACCAGCGCGCGGCTCAACCGTCAACCGCGCGCGCTGGCCAGAACCTGACCGCCACGACC 1192
QY 366 IleThrAspSerAlaThrLysAlaAlaValSerGlnLeuSerAsnLeu---ValGlySer 384
Db 1193 GCCGCTCAAGCGCGGAACAACGTCGCCCGTCGACGGCGCGGCCAACAGTCAACCGTCCG 1252
QY 385 ValGlyValPheAlaGlyTyrTrpThrAlaGlyLeuAlaThrAspProAlaValLysLys 404
Db 1253 ACGGCGTGACCTCGGGCACGACACGCGTGGC-----GCCAAGTCGGCC 1297
QY 405 AlaGluSerPheIleGlnAspLysValLysSerThrAlaSerSerThrThrSerTyrVal 424
Db 1298 GCTTCGGGACCGGTGCTGAGCGTTCGCAACTCGAGCACGACACCAACGCGCGCTATC 1357
QY 425 AlaAspGlnThrValLysLeuAlaLysThrValLysAspMetSerGlyAlaAlaIleSer 444
Db 1358 GCC---GTGACCGGTGGTACGGCCGTGACCGGTGACTCAACGCGCGCAACGCCGTGAAC 1414
QY 445 SerThr 446
Db 1415 ACCACG 1420

RESULT 9

US-09-815-242-4580
; Sequence 4580, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR APPLICATION NUMBER: 2001-03-21
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4580
; LENGTH: 7104
; TYPE: DNA
; ORGANISM: Staphylococcus aureus
US-09-815-242-4580

Alignment Scores:
Pred. No.: 0.00201 Length: 7104
Score: 130.50 Matches: 99
Percent Similarity: 33.20% Conservative: 64
Best Local Similarity: 20.16% Mismatches: 223

Query Match:		5,428	Indels:	105
DB:	10	Gaps:	17	
US-09-825-414-66 (1-487) x US-09-815-242-4580 (1-7104)				
QY	4	AsnGlnSerAlaGlnGlnProProGlyValAlaMetGluSerPheArgThrAlaSer---	22	
DB	2104	AATGCATCATTTACAAGATGAAAAAGATGTAGCAAAATGATAAATGGTAAATTTGAACA	2163	
QY	23	-----AspAlaSerLeuAlaSerSerSerValArgSerValSerThr	36	
DB	2164	AAGCAATTAAAGATATTGTATGCAGCAACAACAATGCACAAGTAGAAGCCATTAAACA	2223	
QY	37	ThrSerCysArgAspLeuGlnAlaIleThrAspTyrLeuLysHisHisValPheAlaAla	56	
DB	2224	AAAGCATCAATGATATTAAATCAAACTGCACCTGCTACAAACAGCTAAAGCAGAGCTCTT	2283	
QY	57	HisArgPheSerValIleGlySerProAspGluArgAspAlaAlaLeu-----Ala	73	
DB	2284	GAAGAATTGACGAAGTTGTTCAAGCACAATTTGATCAAGCACCTTTAAATCCTGATACA	2343	
QY	74	HisAsnGlnGlnIleAspAlaLeuValGluThrArgAlaAsnArgLeuTyrSerGluGly	93	
DB	2344	ACAAATGAGAAGTAGCGGGAAGCTATTGAA-----CGTATTAAATGCAGCTAAA	2391	
QY	94	GluThrProAlaThrIleAlaGluThrPheAlaLysAlaGluLysPheAspArgLeuAla	113	
DB	2392	GTTTCTGGTGTAAAGCAATTGAAAGCCACAACGACTGCACACAGATTTAGAAAGAGTTAA	2451	
QY	114	ThrThrAlaSerSerAlaPheGluAsnThrProPheAlaAlaAlaSerValLeuGlnTyr	133	
DB	2452	AACGAAGAATCTCAAAAATTGAAAT-----ATTACTGACTCTACTCAACA	2499	
QY	134	MetGlnProAlaIleAsnLysGlyAspTyrPheAlaThrProLeuLysProLeuThrPro	153	
DB	2500	AAATGATGCCCTATATATGAAGTTAAACAAGCACAACGCTAGAAAAAATCTCAAAATGCT	2559	
QY	154	LeuIleSerGlyAlaLeuSerGlyAlaMetAspGlnValGlyThrLysMetMetAspArg	173	
DB	2560	ACAGTTTCAATGCAACTAAT-----GAAAGAAGTA-----	2589	
QY	174	AlaArgGlyAspLeuHisTyrLeuSerThrSerProAspLysLeuHisAspAlaMetAla	193	
DB	2590	GCAGAAGCTGATGCAGCAGTAGAAGCAGCTCAAAAAGCAAGTTTACATGCATCCAACTT	2649	
QY	194	ValSerValLysArgHisSerProAlaLeuGlyArgGlnValValAspMetGlyIleAla	213	
DB	2650	GTTAAATCAAAACAGGAAGTTCGTATACAAATCAAAAGTATTAGTAAATCAATGCA	2709	
QY	214	ValGlnThrPheSerAlaLeuAsnValValArgThrValLeuAlaProAlaLeuAlaSer	233	
DB	2710	ATTCAAACA-----	2718	
QY	234	ArgProSerValGlnGlyAlaValAspPheGlyValSerThrAlaGlyGlyLeuValAla	253	
DB	2719	CAAGCAAAAGTTAAACCTGCAGCTGATACGGAAGTAGAAAAACGCA-----	2763	
QY	254	AsnAlaGlyPheGlyAspArgMetLeuSerValGlnSerArgAspGlnLeuArgGlyGly	273	
DB	2764	-----TATATACACGTAAACAAGAAATTCAAAATAGCAAT-----	2799	
QY	274	AlaPheValLeuGlyMetLysAspLysGluProLysAlaAlaLeuSerGluGluThrAsp	293	
DB	2800	-----GCTTCAACTACAGAGAAAGAAACAAGCTGCATATTACAGAA-----	2838	
QY	294	TyrPheAspAlaTyrLysAlaIleLysSerAlaSerTyrSerGlyAlaAlaLeuAsnAla	313	
DB	2839	---TTAGATTAATAAAAGCAAGACAGAACAAATCTTGATGCTGCATAATACAAACAGT	2895	
QY	314	GlyLysArgMetAlaGlyLeuProLeuAspValAlaThr-----AspGlyLeuLys	330	
DB	2896	-----GATGTAAACAACAGCTAAAGACAATGTTATTGCT	2928	

QY	331	AlaValArgSerLeuValSerAlaThrSerLeuThrLysAsnGlyLeuAlaLeuAlaGly	350	
DB	2929	GCAATTAATCAAGTACACAAGCGGCAACAACCTAAGAAATCGGATGCTAAAGCGGAATCGCT	2988	
QY	351	GlyTyrAlaGlyValSerLysLeu-----Gln	359	
DB	2989	CAAAAAGCAAGTGAACGTAATAACTGCCAATTGAAGCAATGAATGATTCGACTACTGAAGAA	3048	
QY	360	LysMetAlaThrLysAsnIleThrAspSerAlaThrLysAlaAlaValSerGlnLeuSer	379	
DB	3049	CAACAAGCTGCAAAAGATAAGTTGATCAAGCAGTAGTACTGCAAAACGCTGATATAGAT	3108	
QY	380	AsnLeuValGlySerValGlyValPheAlaGlyTyrThr-----Ala	394	
DB	3109	AATGCTGCAGCAAAATFACTGATGTAGATAATGCCAAAACTACTAATGAAGCTACATCGCA	3168	
QY	395	GlyLeuAlaThrAspProAlaValLys--LysAlaGluSerPheIleGlnAspLysVal	413	
DB	3169	GCCATTACACCTGATGCAAAATGTTAAACCAACAGCGGAACAACCAATTTGCTGATAAAGTA	3228	
QY	414	LysSerThr-----AlaSerSerThrThrSerTyrVal	424	
DB	3229	CAAGCGCAAGAAACAGCAATTTGATGCTAATTAACGGTGCACACACAGAAAAAGCAGCT	3288	
QY	425	AlaAspGlnThrValLysLeuAlaLysThrValLysAspMetSerGlyGluAlaIleSer	444	
DB	3289	GCGAAACAACAAGTTCAAACTGAAAAACAACAGCTGATACAGCAATTGATGCT--GCA	3345	
QY	445	SerThrGlyAlaSerLeuArgSerThrValAsn	455	
DB	3346	CATACAAATGCAGAAAGTTGAAGCGGCTAAAAAT	3378	

RESULT 10
US-09-815-242-8291
: Sequence 8291, Application US/09815242
: Patent No. US20020061569A1
: GENERAL INFORMATION:
: APPLICANT: Haselbeck, Robert
: APPLICANT: Ohlson, Karl L.
: APPLICANT: Zyskind, Judith W.
: APPLICANT: Wall, Daniel
: APPLICANT: Trawick, John D.
: APPLICANT: Carr, Grant J.
: APPLICANT: Yamamoto, Robert T.
: APPLICANT: Xu, H. Howard
: TITLE OF INVENTION: Identification of Essential Genes in
: TITLE OF INVENTION: Prokaryotes
: FILE REFERENCE: ELITRA.011A
: CURRENT APPLICATION NUMBER: US/09/815,242
: PRIOR FILING DATE: 2001-03-21
: PRIOR APPLICATION NUMBER: 60/191,078
: PRIOR FILING DATE: 2000-03-21
: PRIOR APPLICATION NUMBER: 60/206,848
: PRIOR FILING DATE: 2000-05-23
: PRIOR APPLICATION NUMBER: 60/207,727
: PRIOR FILING DATE: 2000-05-26
: PRIOR APPLICATION NUMBER: 60/242,578
: PRIOR FILING DATE: 2000-10-23
: PRIOR APPLICATION NUMBER: 60/253,625
: PRIOR FILING DATE: 2000-11-27
: PRIOR APPLICATION NUMBER: 60/257,931
: PRIOR FILING DATE: 2000-12-22
: PRIOR APPLICATION NUMBER: 60/269,308
: PRIOR FILING DATE: 2001-02-16
: NUMBER OF SEQ ID NOS: 14110
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 8291
: LENGTH: 7107
: TYPE: DNA
: ORGANISM: Staphylococcus aureus
: FEATURE:
: NAME/KEY: CDS
: LOCATION: (1)...(7107)

US-09-815-242-8291

Alignment Scores:

Pred. No.: 0.00201 Length: 7107
Score: 130.50 Matches: 99
Percent Similarity: 33.20% Conservative: 64
Best Local Similarity: 20.16% Mismatches: 223
Query Match: 5.42% Indels: 105
DB: 10 Gaps: 17

US-09-825-414-66 (1-487) x US-09-815-242-8291 (1-7107)

QY 4 AsnGlnSerAlaGlnGlnProGlyValAlaMetGlnSerPheArgThrAlaSer--- 22
Db 2104 AATGCATCATTTACAGATGAAAAAGATGTAGCAAAATGATAAATTGGTAAATTTGAAACA 2163
QY 23 -----AspAlaSerLeuAlaSerSerSerValArgSerValSerThr 36
Db 2164 AAGGCAATTAAAGATATTGATGCAGCAACAACAATGCACAAGTAGAAGCCATTAAACA 2223
QY 37 ThrSerCysArgAspLeuGlnAlaIleThrAspTyrLeuLysHisHisValPheAlaAla 56
Db 2224 AAAGCGATCAATGATATTATTCAAACTGCACCTGCTACACAGCTAAAGCAGCAGCTCTT 2283
QY 57 HisArgPheSerValIleGlySerProAspGluArgAspAlaAlaLeu-----Ala 73
Db 2284 GAAGAATTTGACGAAGTTGTTCAAGCACAAATGTATCAAGCACCTTTAAATCCGTATACA 2343
QY 74 HisAsnGlnGlnIleAspAlaLeuValGluThrArgAlaAsnArgLeuTyrSerGluGly 93
Db 2344 ACAATGAGAGAGTAGCGGAGCTATTGAA-----CGTATTATGCAGCTAAA 2391
QY 94 GluThrProAlaThrIleAlaGluThrPheAlaLysAlaGluLysPheAspArgLeuAla 113
Db 2392 GTTTCGTGTTAAAGCAATTGAAGCGACACAGCAGCTGCACAAGATTAGAAAGAGTTAAA 2451
QY 114 ThrThrAlaSerSerAlaPheGluAsnThrProPheAlaAlaAlaSerValLeuGlnTyr 133
Db 2452 AACGAAGAAATCTCAAAAATTGAAAT-----ATTACTGACTCTACTCAACA 2499
QY 134 MetGlnProAlaIleAsnLysGlyAspTyrPleuAlaThrProLeuLysProLeuThrPro 153
Db 2500 AAAATGATGCCCTATTAATGAAGTTAAACAAGCAGCAGCAGCTAGAAAAACTCAAAATGCT 2559
QY 154 LeuIleSerGlyAlaLeuSerGlyAlaMetAspGlnValGlyThrLysMetMetAspArg 173
Db 2560 ACAGTTTCAATGCAACTAAT-----GAAGAAGTA----- 2589
QY 174 AlaArgGlyAspLeuHisTyrLeuSerThrSerProAspLysLeuHisAspAlaMetAla 193
Db 2590 GCAGAAAGCTGATGCAGCAGTAGAAGCAGCTCAAAAGCAAGTTTACATGACATCCAAAGTT 2649
QY 194 ValSerValLysArgHisSerProAlaLeuGlyArgGlnValValAspMetGlyIleAla 213
Db 2650 GTTAAATCAAAACAGAACTGCTGATACAAAATCAAAAGTAGATAAATCAATGCA 2709
QY 214 ValGlnThrPheSerAlaLeuAsnValValArgThrValLeuAlaProAlaLeuAlaSer 233
Db 2710 ATTCAACA----- 2718
QY 234 ArgProSerValGlnGlyAlaValAspPheGlyValSerThrAlaGlyGlyLeuValAla 253
Db 2719 CAAGCAAAAGTTAAACCTGCAGCTGATACGGAAGTAGAAACGCA----- 2763
QY 254 AsnAlaGlyPheGlyAspArgMetLeuSerValGlnSerArgAspGlnLeuArgGlyGly 273
Db 2764 -----TATAATACAGCTAAACAAGAAATTCAAATAGCAAT----- 2799
QY 274 AlaPheValLeuGlyMetLysAspLysGluProLysAlaAlaLeuSerGluGluThrAsp 293
Db 2800 -----GCTTCACTACAGAGAAGAAAAACAAGCTGCATATACAGAA----- 2838
QY 294 TripleuAspAlaTyrLysAlaIleLysSerAlaSerTyrSerGlyAlaAlaLeuAsnAla 313

Db 2839 ---TTAGATACTAAAAAGCAAGAAGCAAGAACAAATCTTGATGCTGCATAATACAAACAGT 2895
QY 314 GlyLysArgMetAlaGlyLeuProLeuAspValAlaThr-----AspGlyLeuLys 330
Db 2896 -----GATGTAACAACAGCTAAAGACAATGGTATTGCT 2928
QY 331 AlaValArgSerLeuValSerAlaThrSerLeuThrLysAsnGlyLeuAlaLeuAlaGly 350
Db 2929 GCAATTAAATCAAGTACAAAGCGGCAACAACCTAAGAAATCGGATGCTAAAGCGGAATCGCT 2988
QY 351 GlyTyrAlaGlyValSerLysLeu-----Gln 359
Db 2989 CAAAAAGCAAGTGAACGTTAAACTGCCAATTGAAGCAATGAATGATTGCGACTACTGAAGAA 3048
QY 360 LysMetAlaThrLysAsnIleThrAspSerAlaThrLysAlaAlaValSerGlnLeuSer 379
Db 3049 CAACAAGCTGCAAAAGATTAAGTTGATCAAGCAGTAGTTACTGCAAAACGCTGATATTAGAT 3108
QY 380 AsnLeuValGlySerValGlyValPheAlaGlyTyrThr-----Ala 394
Db 3109 AATGCTGCAGCAAAATACTGATGATATATGCAAAAACCTACTAATGAAGCTACAAATCGCA 3168
QY 395 GlyLeuAlaThrAspProAlaValLys---LysAlaGlnSerPheIleGlnAspLysVal 413
Db 3169 GCCATTACACCTGATGCAAAATGTTAAACCAACAGCGCAACGAATTTGCTGATTAAGTA 3228
QY 414 LysSerThr-----AlaSerSerThrThrSerTyrVal 424
Db 3229 CAAGCGCAAGAAACAGCAATTGATGCTAATTAACGGTGCACAACAGAGAAAAAGCAGCT 3288
QY 425 AlaAspGlnThrValLysLeuAlaLysThrValLysAspMetSerGlyLysAlaIleSer 444
Db 3289 GCGAAACAACAAGTTCAAACTGAAAAAACACACAGCAGCTGATTACAGCAATTGATGGT---GCA 3345
QY 445 SerThrGlyAlaSerLeuArgSerThrValAsn 455
Db 3346 CATACAAATGCAGAAAGTTGAAGCGGCTAAAAAT 3378

RESULT 11

US-09-815-242-4760
; Sequence 4760, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 4760
LENGTH: 6228
TYPE: DNA
ORGANISM: Staphylococcus aureus
US-09-815-242-4760

Alignment Scores:
Pred. No.: 0.00212 Length: 6228
Score: 129.50 Matches: 106
Percent Similarity: 35.88% Conservative: 91
Best Local Similarity: 19.31% Mismatches: 201
Query Match: 5.38% Indels: 151
DB: 10 Gaps: 22

US-09-825-414-66 (1-487) x US-09-815-242-4760 (1-6228)

QY 20 ThrAlaSerAspAlaSerLeu-----AlaSerSerSerValArgSerValSer 35
DB 2788 ACTGCAGCAACACAAACATTTTAAACGCTAATACCAATGCTGATGTTGAACAAGTAAAG 2847
QY 36 ThrThrSerCysArgAspLeuGlnAlaIleThrAspTyrLeuLysHisValPheAla 55
DB 2848 ACAAAATGCCATTCAAGGAATACAAAGCAATTACACCAGCTACAAA-----2892
QY 56 AlaHisArgPheSerValIleGlySerProAspGluArgAspAlaAlaLeuAlaHisAsn 75
DB 2893 -----GTAAAAACAGATGCMAAAATGCC-----2916
QY 76 GluGlnIleAspAlaLeuValGluThrArgAlaAsnArgLeuTyrSerGluGlyGluThr 95
DB 2917 -----ATCGATAAAGTGCAGAAACGACATATATGATATTTAATATATATGAT--- 2967
QY 96 ProAlaThrIleAlaGluThrPheAlaLysAlaGluLysPheAspArgLeuAlaThrThr 115
DB 2968 ---GCCAGCGCTCGAAGAACACACAGCAGCACAATTAATCTGATCAAGCTGTAGCCACA 3024
QY 116 AlaSerSerAlaPheGluAsnThrProPheAlaAlaAlaSerValLeuGlnTyrMetGln 135
DB 3025 GCGAAGCAAAATATTAAT-----GCAGCAGATACGAATCAAGAAAGTTGCA 3069
QY 136 ProAlaIleAsnLysGlyAspTyrPleuAlaThrProLeuLysProLeuThrProLeuIle 155
DB 3070 CAAGCAAAAGATCAGGGCACACAAATATAGTAGTCAACCGGCCACA-----3120
QY 156 SerGlyAlaLeuSerGlyAlaMetAspGlnValGlyThrLys-----MetMet 171
DB 3121 -----CAAGTTAAACGCGATACTCGCAATGTTGTAAT 3153
QY 172 AspArgAlaArgGlyAspLeuHisTyrLeuSerThrSerProAspLysLeuHisAspAla 191
DB 3154 GATAAAGCGCGAGGCGATTAACAATATCATGCTACAACTGGCGGACTCGAGAAGAG 3213
QY 192 MetAlaValSerValLysArgHisSerProAlaLeuGlyArgGlnValValAspMetGly 211
DB 3214 AAACAAGAGCGATTAATCGTGTCAATACACTTAAAAATAGAGCATTAAGTATTTGGT 3273
QY 212 IleAlaValGlnThrPheSerAlaLeuAsnValValArgThr-----225
DB 3274 GTG---ACGTTACTACTGCGATGTCATAGTATTAGAGAGCATGCAATCAATCAATC 3330
QY 226 ---ValLeuAlaProAlaLeuAlaSerArgProSerValGlnGlyAla-----240
DB 3331 GCGCAGTTCAACCGCATGTAAAGAAACAAACTGCTACAGGTGTATTAAATGATTTA 3390
QY 240 -----240
DB 3391 GCAACTGCTAAAAAGCAAGAAATTAATCAAAACAAATGCACAACTGAAGAAAGCAA 3450
QY 241 -----ValAspPheGlyValSerThrAlaGlyGlyLeuValAlaAsnAla 255
DB 3451 GTGGCTTTAAATCAAGTGATCAAGAGATTACCAACGGCA-----ATTAAATATATA 3501
QY 256 GlyPheGlyAspArgMetLeuSerValGlnSerArgAspGlnLeuArgGlyGlyAlaPhe 275

DB 3502 AATCAAGCTGATACAAATATGCGGAAGTAGATCAAGCGCAACAATATAGTACAAAGCAATT 3561
QY 276 -----ValLeuGlyMetLysAspLysGlu 283
DB 3562 AATCGGATTACGCCCAATATTTGTTAAAAAACCTGCAGCATTAGCAACAATCATCAGCAT 3621
QY 284 ProLysAlaAlaLeuSerGluGluThrAspTyrPleuAspAlaIleTyrLysSer 303
DB 3622 TATAATGCTAAATTAGCTGAATCAATGCTACACCAGATGCAACCAATGATGAGAAAAAT 3681
QY 304 AlaSerTyrSerGlyAlaAlaLeuAsnAlaGlyLysArgMetAla-----318
DB 3682 GCTCGATCAAT-----ACTTTAATCAAGACACAGACACAAAGCTATTGAAAGTATTAA 3735
QY 319 -----GlyLeuProLeuAsp-----ValAlaThrAspGlyLeuLys 330
DB 3736 CAAGCTAACACAAATGCAGAGTAGACCAAGCTGCAGACAGTAGCAGAGAAATATATCGAT 3795
QY 331 AlaValArgSerLeuValSerAlaThrSerLeuThrLysAsnGlyLeuAlaLeuAlaGly 350
DB 3796 GCTGTTCAAGTTGATGTAGTAAAAAAACAAGCAGCGGAGATTAATACTACT-----3846
QY 351 GlyTyrAlaGlyValSerLysLeuGlnLysMetAla-----ThrLysAsnIleThrAsp 368
DB 3847 -----GCTGAAGTGGCGAAGCGTATGGAAGCGGTTAACAACAACCTAATGCACTGAC 3900
QY 369 SerAlaThrLysAlaAlaValSerGlnLeuSerAsnLeuValGlySerValGlyValPhe 388
DB 3901 GAAGAAAGCAGCGCTGCTGTTAATCAATCAACTT-----3939
QY 389 AlaGlyTyrThrThrAlaGlyLeuAlaThrAspProAlaValLysLysAla---GluSer 407
DB 3940 -----AAAGATCAAGCAATTAATCAAAATTAATCAAAAC 3972
QY 408 PheIleGlnAspLysValLysSerThrAlaSerSerThrThrSerTyrVal-----424
DB 3973 CAACCAAAATGATCAGGTAGACACAACATCAAAATCAAGCGGTAATGCTATAGATAATGTT 4032
QY 425 ---AlaAspGlnThrValLys-----LeuAlaLysThrValLysAsp 437
DB 4033 GAAGCTGAAGTAGTAATTAACCAAGGCAATTGCAGATATGAAAAAAGCTGTTAAAGAA 4092
QY 438 MetSerGlyLysAlaIleSerSerThrGlyAlaSerLeuArgSerThrValAsnAsnLeu 457
DB 4093 AAGCAACAGCAATTGATAAT-----AGTCTGATTCACACAGATAATGAGAAA 4140
QY 458 ArgHisArgSerAlaProGluAlaAspIleGluGlyGlyIleSerAlaPheSerArg 477
DB 4141 GAAGTTCCTTCACAAGCATTAGCTTAAGAAAAAGAAAGACACTTCAGCTATTGACCAG 4200
QY 478 SerGluThrProPheGlnLeuArgArg 486
DB 4201 GCTCAACGAATAGTCAGGTGAATCAA 4227
RESULT 12
US-09-815-242-8815
Sequence 8815, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haselbeck, Robert
APPLICANT: Ohlsen, Kari L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815,242
CURRENT FILING DATE: 2001-03-21


```
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8815
; LENGTH: 6561
; TYPE: DNA
; ORGANISM: Staphylococcus aureus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(6561)
US-09-815-242-8815
```

Alignment Scores:

Pred. No.:	0.00228	Length:	6561
Score:	129.50	Matches:	106
Percent Similarity:	35.88%	Conservative:	91
Best Local Similarity:	19.31%	Mismatches:	201
Query Match:	5.38%	Indels:	151
DB:	10	Gaps:	22

US-09-825-414-66 (1-487) x US-09-815-242-8815 (1-6561)

```
QY 20 ThrAlaSerAspAlaSerLeu-----AlaSerSerSerValArgSerValSer 35
Db 2788 ACTGCAGCAACAACACATTTTAAACGCTAATACCAATGCTGATGTTGAACAAGTAAAG
QY 36 ThrThrSerCysArgAspLeuGlnAlaIleThrAspTyrLeuLysHisHisValPheAla 55
Db 2848 ACAAAATGCCATTCAAGGAATACAAACAATTACACCAGCTACAAA-----
QY 56 AlaHisArgPheSerValIleGlySerProAspGluArgAspAlaAlaLeuAlaHisAsn 75
Db 2893 -----GTA AAAACAGATGCAAAAAATGCC-----
QY 76 GluGlnIleAspAlaLeuValGluThrArgAlaAsnArgLeuTyrSerGluGluGluThr 95
Db 2917 -----ATCGATAAAAGTCGGGAACGCAACATATACGATTAATTAATATGAT---
QY 96 ProAlaThrIleAlaGluThrPheAlaLysAlaGluLysPheAspArgLeuAlaThrThr 115
Db 2968 ---GCGACGCTCGAAGAACAACAAGCAGCACAACAATTACTTGATCAAGCTGTAGCCACA 3024
QY 116 AlaSerSerAlaPheGluAsnThrProPheAlaAlaAlaSerValLeuGlnTyrMetGln 135
Db 3025 GCGAAGCAAAATATTAT-----GCAGCAGATACGAATCAAGAGAGTTGCA 3069
QY 136 ProAlaIleAsnLysGlyAspTrpLeuAlaThrProLeuLysProLeuThrProLeuIle 155
Db 3070 CAAGCAAAAAGATCAGGGCACAACAATAATAGTAGTGAATCAACCGGCAACA----- 3120
QY 156 SerGlyAlaLeuSerGlyAlaMetAspGlnValGlyThrLys-----MetMet 171
Db 3121 -----CAAGTTAAACGCGTACTCGCAATGTTGTAAT 3153
QY 172 AspArgAlaArgGlyAspLeuHisTyrLeuSerThrSerProAspLysLeuHisAspAla 191
Db 3154 GATAAAGCGCGAGAGGCGATACAAATATCAATGCTACAACTGGCGGCACTGAGAGAG 3213
QY 192 MetAlaValSerValLysArgHisSerProAlaLeuGlyArgGlnValValAspMetGly 211
Db -----
```

```
Db 3214 AAACAAGAAGCGATAAATCGTGTCAATACACTTAAAAAATAGAGCATTTAACTGATATTTGGT
QY 212 IleAlaValGlnThrPheSerAlaLeuAsnValValArgThr-----
Db 3274 GTG---ACGTCTACTACTGCGATGCTCAATAGTATTAGAGACGATGCATCAATAATC 3330
QY 226 ---ValLeuAlaProAlaLeuAlaSerArgProSerValGlnGlyAla----- 240
Db 3331 GGGCGAGTTCAACCGCATGTAGAGAAACAACAACTGCTACAGGTGTTAATGATTTA 3390
QY 240 ----- 240
Db 3391 GCAACTGCTAAAAAGCAAGAAATTATCAAAAACACAATGCAACACTGAAGAAAGCAA 3450
QY 241 -----ValAspPheGlyValSerThrAlaGlyGlyLeuValAlaAsnAla 255
Db 3451 GTGGCTTTAAATCAAGTGATCAAGAGTTAGCAACGGCA-----ATTAATTAATTA 3501
QY 256 GlyPheGlyAspArgMetLeuSerValGlnSerArgAspGlnLeuArgGlyAlaPhe 275
Db 3502 AATCAAGCTGATACAATGCGGAAGTAGATCAAGCGCAACAATTAGGTACAAAAGCAATT 3561
QY 276 -----ValLeuGlyMetLysAspLysGlu 283
Db 3562 AATGCGATTCAGCCAAATATTGTTAAAAAACCTGCAGCATTTAGCACAATCAATCAGCAT 3621
QY 284 ProLysAlaAlaLeuSerGluGluThrAspTrpLeuAspAlaTyrLysAlaIleLysSer 303
Db 3622 TATAATGCTTAATTAAGTGAATCAATGCTACACAGATGCAACGAATGATGAGAAAAT 3681
QY 304 AlaSerTyrSerGlyAlaAlaLeuAsnAlaGlyLysArgMetAla----- 318
Db 3682 GCTGCGATCAAT-----ACTTTAAATCAAGACAGACAGACAAGCTTATTGAAGTATTAAA 3735
QY 319 -----GlyLeuProLeuAsp-----ValAlaThrAspGlyLeuLys 330
Db 3736 CAAGCTAACACAATAATGCAGAAAGTAGACCAAGCTGCAGACAGACAGAGATATATCGAT 3795
QY 331 AlaValArgSerLeuValSerAlaThrSerLeuThrLysAsnGlyLeuAlaLeuAlaGly 350
Db 3796 GCTGTTCAAGTTGATGTAGTAAAAAAACAAGCAGCGCGAGATTAATAATCACT----- 3846
QY 351 GlyTyrAlaGlyValSerLysLeuGlnLysMetAla-----ThrLysAsnIleThrAsp 368
Db 3847 -----GCTGAAGTGGCGAAGCGTATTGAAGCGGTTAAACAACAACCTAATGCAACTGAC 3900
QY 369 SerAlaThrLysAlaAlaValSerGlnLeuSerAsnLeuValGlySerValGlyValPhe 388
Db 3901 GAAGAAAAGCAGGCTGCTGTTAATCAAAATCAATCAACTT----- 3939
QY 389 AlaGlyTyrThrThrAlaGlyLeuAlaThrAspProAlaValLysLysAla---GluSer 407
Db 3940 -----AAAGATCAAGCAATTAAATCAAAATTAAATCAAAAC 3972
QY 408 PheIleGlnAspLysValLysSerThrAlaSerSerThrThrSerTyrVal----- 424
Db 3973 CAACAATAATGATCAGGTAGACACAACATAACAATCAAGCGGTAAATGCTATAGATAATGTT 4032
QY 425 ---AlaAspGlnThrValLys-----LeuAlaLysThrValLysAsp 437
Db 4033 GAAGCTGAAGTAGTAATTAAACCAAGGCAATTGCAAGATATTGAAAAAGCTGTTAAAGAA 4092
QY 438 MetSerGlyGluAlaIleSerSerThrGlyAlaSerLeuArgSerThrValAsnAsnLeu 457
Db 4093 AAGCAACAGCAAAATTGATTAAT-----AGTCTGATTTCAACAGATTAATGAGAAA 4140
QY 458 ArgHisArgSerAlaProGluAlaAspIleGluGluGlyIleSerAlaPheSerArg 477
Db 4141 GAAGTGTCTTACAAAGCATTAAGCTAAAGAAAAAGAAAAAGCACTTGACGATATTGACCAA 4200
QY 478 SerGluThrProPheGlnLeuArgArg 486
Db 4201 GCTCAAAACGAATAGTCAGGTGAATCAA 4227
```

```
RESULT 13
US-09-815-242-8615
; Sequence 8615, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8615
; LENGTH: 7035
; TYPE: DNA
; ORGANISM: Staphylococcus aureus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(7035)
US-09-815-242-8615

Alignment Scores:
Pred. No.: 0.00283 Length: 7035
Score: 129.00 Matches: 96
Percent Similarity: 38.52% Conservative: 97
Best Local Similarity: 19.16% Mismatches: 231
Query Match: 5.36% Indels: 79
DB: 10 Gaps: 12

US-09-825-414-66 (1-487) x US-09-815-242-8615 (1-7035)
QY 4 AsnglnSerAlaGlnGlnProGlyValAlaMetGlnSerPheArgThrAlaSerasp 23
DB 4204 TCAAACTCAGCAAGCAAGTGCCTCAACCTCAACAAGTACAAAGCAGCGAATCCGAT 4263
QY 24 AlaSerLeuAlaSer-----SerSerValArgSerValSerThrThrSer 38
DB 4264 TCACAAAGCACATCATATATACAAAGTACATCAACCAAGCCAAAGTGAATCCACATCGACA 4323
QY 39 CysArgAspLeuGlnAlaIleThrAsp-TyrLeuLys-----HisHisValPheAl 55
DB 4324 TCAACGTCACCTAAGCGATTCACAAAGTATATCTAAAGTACAAAGCCAAATCAGGTTCAACA 4383
QY 55 aAlaHisArgPheSerValIleGlySerProAspGluArgAspAlaAlaLeuAlaHisAs 75
DB 4384 AGCACATCT-GCATCATTAAGTGGTTCAGAAAGCGAATCCGATTCACAAAGCATATCAAC 4442
QY 75 nGluGlnIleAspAlaLeuValGluThrArgAlaAsnArgLeuTyrSerGluGluTh 95
DB 4443 AAGCACAAGTGAAGTCAAAATCAGAAAGTACATCCACATCGCTCAGTGAATCAACAAGTAC 4502
```

```
QY 95 r-----ProAlaThrIleAlaGluThrPheAlaLysAlaGlu 108
DB 4503 AAGCAACTCAGATCAGCAAGTACGTCACACATGCTCCAGTAAGTACAGCAAGCGCAAGTGA 4562
QY 108 spHeAspArgLeuAlaThrThrAlaSerSerAlaPheGluAsnThrProPheAlaAla 128
DB 4563 ATCCGATTCGTCGTCACACATCATTAAGTATTCACGTCAGCGTCATGCAAGCAGTGA 4622
QY 128 aSerValLeuGlnIleTyrMetGlnProAlaIleAsnLysGlyAspTrpLeuAlaThrPro 148
DB 4623 ATCCGATTCACAAAGCACATCAACGTCATTAAGTAAATTCCAAAGTACATCAACATCAAT 4682
QY 148 ulysProLeuThrProLeuIleSerGlyAlaLeuSerGlyAlaMetAspGlnValGlyTh 168
DB 4683 TCGA--ATGTCGACTATTGCAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 4739
QY 168 rLysMetMetAspArgAlaArgGlyAspLeuHisTyrLeuSerThrSerProAsp-LysL 188
DB 4740 AACATCAGAAAGTACAGTGAATCCGATTCGACATCAACATCATTAAGCATTCACAAAG 4799
QY 188 euHisAspAlaMetAlaValSerValLysArg----- 198
DB 4800 CACATCAAGAAGTACAGTGCATCTGGATCCGCAAGTACATCAACATCAACAAGTGACTC 4859
QY 199 -----HisSerProAlaLeuGlyArgGlnValValAspMetGlyIleAlaValGln 216
DB 4860 TCGTAGTACATCAGCTTCAACTAGTACTTCGATCGGTACAAAGTACCGCTGATTCA-CAAA 4918
QY 216 hrPheSerAlaLeuAsnValValArgThrValLeuAlaProAlaLeuAlaSerArgPro 236
DB 4919 GTATGTCGCTTCAACAAGTACATCAACAGTGAAGTGAATTCACACGTCATTATCTGATA 4978
QY 236 erValGlnGlyAlaVal-----AspPheGlyValSerThrAlaGlyLeuValAlaA 254
DB 4979 GTGTGAGTGAATTCACATCAGACTCAACAAGTGAAGTGAATCTGTTGAGTGAAGTGCCT 5038
QY 254 snAlaGlyPheGlyAspArgMetLeuSerValGlnSerArgAspGlnLeuArgGlyGlyA 274
DB 5039 CTATATCGTTAAGTGAATTCGACAAAGCACATCAACATCGGCTAGTGAAGTAAATG- 5091
QY 274 lApheValLeuGlyMetLysAspLysGluProLysAlaAlaLeuSerGluGluThrAsp 294
DB 5092 -----ACCGCAAGCATATCTGATTCCAAAGTA 5119
QY 294 rPLeuAspAlaTyrLysAlaIleLysSerAlaSerTyrSerGlyAlaAlaLeuAsnAlaG 314
DB 5120 TGTCAAGATCTGTAATGATTTCAGAAAGTGAAGTGAATCTAATTCTGAAAGTGACTCA- 5178
QY 314 lYlYsArgMetAlaGlyLeuProLeuAspValAlaThrAspGlyLeuLysAlaValArg 334
DB 5179 --AAATCGATGAGTGT--TCAACAAGCGTAAGTGAATCTGGC-----T 5218
QY 334 erLeuValSerAlaThrSerLeuThrLysAsnGlyLeuAlaLeuAlaGlyGlyTyrAlaG 354
DB 5219 CATTTGAGCGTCTCAACGTCATTAAAGAAA----- 5247
QY 354 lYValSerLysLeuGlnLysMetAlaThrLysAsnIleThrAspSerAlaThrLysAlaA 374
DB 5248 -----TCAGAAAGTGTAAGCGAGTCTATTTCATTGAGTG 5281
QY 374 lAlaSerGlnLeuSerAsnLeuValGlySerValGlyValPheAlaGlyTyrThrThra 394
DB 5282 GCTCACAAATCGATGAGTGAATTCAGTAAGCACAAAG-----GATTGCT 5323
QY 394 lAglyLeuAlaThrAspProAlaValLysLysAlaGluSerPheIleGlnAspLysValL 414
DB 5324 CATCATTAAGTGAATTCGACGTCACAAAGAGTTCAGAAAGCGTGAATCTGATTGATCAT 5383
QY 414 ysSerThrAlaSerSerThrThrSerTyrValAlaAspGlnThrValLysLeuAlaLys 434
DB 5384 TAAAGTATTCAAATCAACAAGTGTTCGACTTCACAACAAGTACATCAGTTCATTGAGCA 5443
```


QY 434 hrValLysAspMetSerGlyGluAlaIleSerSerThrGlyAlaSerLeuArgSerThrV 454
|| ::||| ||| ::||| ::|
Db 5444 CTTCGACATCATTAAGTGTTCAGAAAGTCTAAGCGAGTCTAGCTCGCTAAGTATCA 5503
QY 454 aLAsnAsnLeuArgHisArgSerAlaProGluAlaAspIleGluGlyIleSer 473
::: ||| ::::||| ||| |||
Db 5504 TATCAATGATGATTTCTACTAGTACAAGTGACTCCGACTCATTAAGTGCATCAATATCT 5562

RESULT 14

US-09-815-242-4761
; Sequence 4761, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4761
; LENGTH: 7434
; TYPE: DNA
; ORGANISM: Staphylococcus aureus
US-09-815-242-4761

Alignment Scores:
Pred. No.: 0.00438 Length: 7434
Score: 127.50 Matches: 106
Percent Similarity: 31.77% Conservative: 70
Best Local Similarity: 19.13% Mismatches: 235
Query Match: 5.30% Indels: 143
DB: 10 Gaps: 18

US-09-825-414-66 (1-487) x US-09-815-242-4761 (1-7434)

QY 4 AsnGlnSerAlaGlnGlnProProGlyValAlaMetGluSerPheArgThrAlaSer--- 22
||| ||| ||| ::|
Db 2116 AATGATCATCATTAACAAGATGAAAAAGATGTAGCAAAATGATAAAATTTGTAATAATTGAACA 2175
QY 23 -----AspAlaSerLeuAlaSerSerSerValArgSerValSerThr 36
||| ||| ||| ::|
Db 2176 AAGCAATTAAGATATTTGATGACAGCAACAACAATGACACAAGTAGAAGCCATTAAACA 2235
QY 37 ThrSerCysArgAspLeuGlnAlaIleThrAspTyrLeuLysHisValPheAlaAla 56
::: ||| ::|
Db 2236 AAAGCAATCATGATATTAATCAAACTACACTGCTACACACAGCTAAAGCAGCAGCTCTT 2295
QY 57 HisArgPheSerValIleGlySerProAspGluArgAspAlaAlaLeu-----Ala 73
||| ::| ||| |||
Db 2296 GAAGAATTTGACGAAGTTGTTCAAGCACACAAATTGATCAAGCACCTTTAAATCCTGTATACA 2355

QY 74 HisAsnGluGlnIleAspAlaLeuValGluThrArgAlaAsnArgLeuTyrSerGluGly 93
||| ||| ::| ||| ||| ::|
Db 2356 ACAATGAAGAGTAGCGGAAGCTATTGAA-----CGATTAAATGCAAGCTAAA 2403
QY 94 GluThrProAlaThrIleAlaGluThrPheAlaLysAlaGluLysPheAspArgLeuAla 113
::: ||| ||| ::|
Db 2404 GTTCTGTGTAAAGCAATTGAAGCGACAACGACTGCACAAGATTAGAAAGAGTTAAA 2463
QY 114 ThrThrAlaSerSerAlaPheGluAsnThrProPheAlaAlaAlaSerValLeuGlnTyr 133
||| ||| |||
Db 2464 AACGAAGAATCTCAAAAATTGAAAT-----ATTACTGACTCTACGCAACA 2511
QY 134 MetGlnProAlaIleAsnLysGlyAspTyrPheAlaThrProLeuLysProLeuThrPro 153
||| ||| ::| ||| |||
Db 2512 AAATGGATGCCCTTAATGAAGTTAAACAAGCTGCAACAGCTAGAAAAGCTCAAAATGCT 2571
QY 154 LeuIleSerGlyAlaLeuSerGlyAlaMetAspGlnValGlyThrLysMetMetAspArg 173
::: ||| ||| ::|
Db 2572 ACAGTTCAAAATGCAACAAT-----GAAGAAGTA----- 2601
QY 174 AlaArgGlyAspLeuHisTyrLeuSerThrSerProAspLysLeuHisAspAlaMetAla 193
||| ||| ::| ||| ||| |||
Db 2602 GCAGAACTGATGCAGCAGTAGATGCAGCTCAAAAGCAAGTTTACATGACATCCAAAGTT 2661
QY 194 ValSerValLysArgHisSerProAlaLeuGlyArgGlnValAlaAspMetGlyTleAla 213
||| ||| ::| ||| ||| |||
Db 2662 GTTAAATCAAAACGGAAGTTGCTGATACAAAATCAAAAGTATTAGATAAATCAATGCA 2721
QY 214 ValGlnThrPheSerAlaLeuAsnValValArgThrValLeuAlaProAlaLeuAlaSer 233
::: ||| |||
Db 2722 ATTCAACA----- 2730
QY 234 ArgProSerValGlnGlyAlaValaAspPheGlyValSerThrAlaGlyGlyLeuValAla 253
::: ||| ::| ||| ||| |||
Db 2731 CAAGCAAAAGTTAAACCTGCAGCTGATACGGAAGTAGAAACGCA----- 2775
QY 254 AsnAlaGlyPheGlyAspArgMetLeuSerValGlnSerArgAspGlnLeuArgGlyGly 273
::: ||| ||| ::|
Db 2776 -----TATATACACGTTAACACAGAATTCAAAATAGCAAT----- 2811
QY 274 AlaPheValLeuGlyMetLysAspLysGluProLysAlaAlaLeuSerGluGluThrAsp 293
::: ||| ||| ||| |||
Db 2812 -----GCTTCAACTACAGAAAGAAAACAAGCTGCATATACAGAA----- 2850
QY 294 TrpLeuAspAlaTyrLysAlaIleLysSerAlaSerTyrSerGlyAlaAlaLeuAsnAla 313
||| ||| ||| ::| ||| |||
Db 2851 ---TTAGATACTTAAAAAGCAAGCAAGCAACAATCTTGATGCTGCAAAATACAAACAGT 2907
QY 314 GlyLysArgMetAlaGlyLeuProLeuAspValAlaThr-----AspGlyLeuLys 330
||| ||| ||| |||
Db 2908 -----GATGTAAACAACAGCTAAAGACAAATGATTGCT 2940
QY 331 AlaValArgSerLeuValSerAlaThrSerLeuThrLysAsnGlyLeuAlaLeuAlaGly 350
||| ::| ::| ||| ||| |||
Db 2941 GCAATTAAATCAAGTTCAAGCTGCCACACTTAACAATCGGATGCAAAAGCGCAAAATCGCT 3000
QY 351 GlyTyrAlaGlyValSerLysLeu-----Gln 359
||| ||| |||
Db 3001 CAAAAAGCAAGTGAAGCTTAAACAGCAATTGCAAGCAATGAATGATTCGACTACTGAAGAA 3060
QY 360 LysMetAlaThrLysAsnIleThrAspSerAlaThrLysAlaAlaValSerGlnLeuSer 379
::: ||| ||| ::| ||| ||| |||
Db 3061 CAACAAGCAGCGAAGACAAAGTGTGATCAAGCAGTAGTTACTGCAAAACGCTGATATAGAT 3120
QY 380 AsnLeuValGlySerValGlyValPheAlaGlyTyrThr-----Ala 394
||| ::| ||| ||| |||
Db 3121 AATGCTGCAGCAACAATGATGTGATTAATGCAAAAACACTACAATGAAGCTACAAATCGCA 3180
QY 395 GlyLeuAlaThrAspProAlaValLys---LysAlaGluSerPheIleGlnAspLysVal 413
::: ||| ||| ||| ||| ||| |||
Db 3181 GCCATTACACCTGATGCAAAATGTTAAACCAAGCAGCAAAACAAAGCAATTCAGATTAAGATA 3240

[illegible]

DB:	10	Gaps:	18
US-09-825-	414-66 (1-487) x US-09-815-242-8869 (1-7437)		
QY	4 AsngInSerAlaGlnGlnProProGlyValAlaMetCluSerPheArgThrAlaSer---		22
Db	2116 AATGCATCATTAACAAGATAAAGATGTAGCAAAATGATAAAATTGGTAAAAATTGAACA		2175
QY	23 -----AsPalaserLeuAlaSerSerSerValArgSerValSerThr		36
Db	2176 AAGC CATTAAGA GATATTGATGCAGCACACACAACAAATGCACAAGTAGAAGCCATTA AAAACA		2235
QY	37 ThrSerCyArGaSpLeuGlnAlaIleThrAspTyrLeuLysHisHisValPheaIaAla		56
Db	2236 AAAGCAATCAATGATATTATTCACAACTACACCCTGCTACACACAGCTAAGCAGCAGCTCTT		2295
QY	57 HisArgPheSerValIleGlySerProAspGluArgAspAlaAlaLeu-----Ala		73
Db	2296 GAAGAA TTTGACCGAAGTGTTCTCAAGCACACAAATTGATCAAGCACACCTTTAAATCC TGATACA		2355
QY	74 HisAsnGlnGlnIleAspAlaLeuValGluThrArgAlaAsnArgLeuTyrSergGluGly		93
Db	2356 ACAAAATGAGAAGATAGCGGAAGCTATTGAA-----CGTATTAAATGCAGCTAAA		2403

```

RESULT 15
US-09-815-242-8869
; Sequence 8869, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8869
; LENGTH: 7437
; TYPE: DNA
; ORGANISM: Staphylococcus aureus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(7437)
US-09-815-242-8869

```

QY	94	GIUThrProAlaThrIleAlaGluThrPheAlaLysAlaGluLysPheAspArgLeuAla	113
Db	2404	GTTTCTGTTGTTAAAGCAATTGAAGCGACACGACTGCACAAGATTTAGAAAGAGTTAAA	2463
QY	114	ThrThrAlaSerSerAlaPheGluAsnThrProPheAlaAlaAlaSerValLeuGlnTyr	133
Db	2464	AACGAAGAAATCTCAAAAATTGAAAAT-----ATTACTGACTCTACGCCAACA	2511
QY	134	MetGlnProAlaIleAsnLysGlyAspTrpLeuAlaThrProLeuLysProLeuThrPro	153
Db	2512	AAAATGATGCCATTATAATGAAGTTAAACAAGCTGCACAAGCTAGAAAAAGCTCAAAATGCT	2571
QY	154	LeuIleSerGlyAlaLeuSerGlyAlaMetAspGlnValGlyThrLysMetMetAspArg	173
Db	2572	ACAGTTTCAAAATGCAACAAT-----GAAGAGTA-----	2601
QY	174	AlaArgGlyAspLeuHisTyrLeuSerThrSerProAspLysLeuHisAspAlaMetAla	193
Db	2602	GCAGAAGCTGATGCAGCAGTAGATGCAGCTCAAAAACAAGGTTTACATGACATCCAAAGTT	2661
QY	194	ValSerValLysArgHisSerProAlaLeuGlyArgGlnValValAspMetGlyIleAla	213
Db	2662	GTTAAATCAAAAACAGCAAGTTGCTGATACAAAATCAAAAAGTATTAGATAAAATCAATGCA	2721
QY	214	ValGlnThrPheSerAlaLeuAsnValValArgThrValLeuAlaProAlaLeuAlaSer	233
Db	2722	ATTCAACA-----	2730
QY	234	ArgProSerValGlnGlyAlaValAspPheGlyValSerThrAlaGlyGlyLeuValAla	253
Db	2731	CAAGCAAAAAGTTAAACCTGCACAGCTGATACGGAAGTAGAAAAACGCA-----	2775
QY	254	AsnAlaGlyPheGlyAspArgMetLeuSerValGlnSerArgAspGlnLeuArgGlyGly	273
Db	2776	-----TATAATACACGTAACAAGAATAATCCAATAATGACAAT-----	2811
QY	274	AlaPheValLeuGlyMetLysAspLysGluProLysAlaAlaLeuSerGluGluThrAsp	293
Db	2812	-----GCTTCACTACAGAGAAGAAAAACAAGCTGCATATACAGAA-----	2850
QY	294	TrpLeuAspAlaTyrLysAlaIleLysSerAlaSerTyrSerGlyAlaAlaLeuAsnAla	313

Alignment Scores:	
Pred. No.:	0.00438
Score:	127.50
Percent Similarity:	31.77%
Best Local Similarity:	19.13%
Query Match:	5.30%
Length:	7437
Matches:	106
Conservative:	70
Mismatches:	235
Indels:	143

QY	314	GLYLYSARGMETALAGLYLEUPROLEUASPV	ALATHR-----	ASPGLYLEU	LYS	330
					:::	:::
Db	2908	-----	GATGTACACACAGCTAAAGAC	ATACTATTGCT		2940
QY	331	ALAVALARQSERLEUVALSERALATHRSER	LEUTHRLYSASNGLYLEU	ALAEU	ALAGLY	350


```

Db 2941 GCAATTAAATCAAGTTCAAGCTGCCACACTAAGAAATCGGATGCAAAAAGCGAAATCGCT 3000
      |||::: :::: |||::: |||
QY 351 GlyTyrAlaGlyValSerLysLeu-----Gln 359
      ||| |||
Db 3001 CAAAAAGCAAGTGAAGCTAAACACGCAATTGAACCAATGATTCGACTACTGAAGAA 3060
      ||| |||
QY 360 LysMetAlaThrLysAsnIleThrAspSerAlaThrLysAlaAlaValSerGlnLeuSer 379
      :::: |||::: ||| ||| ||| ::::
Db 3061 CAACAAGCAGCAGCAAGACAAAGTGATCAAGCAGTAGTACTGCAAAACGCTGATATAGAT 3120
      ||| |||
QY 380 AsnLeuValGlySerValGlyValPheAlaGlyTrpThr-----Ala 394
      ||| :::: ||| ||| ||| |||
Db 3121 AATGCTGACGCAACCAATGATGTGATTAATGCAAAACTACAAATGAAGCTACATCGCA 3180
      ||| ||| ||| ||| ||| |||
QY 395 GlyLeuAlaThrAspProAlaValLys--LysAlaGlnSerPheIleGlnAspLysVal 413
      :::: ||| ||| ||| ||| ||| ||| |||
Db 3181 GCCATTACACCTGATGCAATGTGTTAAACCAGCAGCAAAACAAGCAATTGCAGATAAGTA 3240
      ||| ||| ||| ||| ||| ||| |||
QY 414 LysSer----- 415
      ::::
Db 3241 CAAGCTCAAGAAACAGCAATTGATGAAATAACGGCTCAACAACCTGAAGAAACAGAGCT 3300
      ||| |||
QY 416 -----ThrAlaSerSerThrThrSerTyrValAlaAsp----- 426
      ||| |||
Db 3301 GCTAAACACCAAGTCAACTGAAAAACCAACAGCTGATGCCGCAATGATGACACACAT 3360
      ||| |||
QY 427 -----GlnThrValLys 430
      ||| |||
Db 3361 ACAAAATGCGAAGTTGAAGCGGCTAAAAAAGCAGCAATTGCTAAATGGAAGCGATTCAG 3420
      ||| ||| ||| |||
QY 431 LeuAlaLysThrValLysAspMetSerGlyGluAlaIleSerSerThrGlyAlaSerLeu 450
      ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3421 CCAGCAACACCAACTAAAGATAATGCGAAGAAGCAATTGCTACGAAAGCGAATGAACGT 3480
      ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 451 ArgSerThrValAsnAsnLeuArgHisArgSerAlaProGlu-----Ala 465
      :::: :::: :::: ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3481 AAAACAGCAATCGCTCAACGCAAGACATTTACTGCTGAAGAATAATGCAGCGGCTAATGCG 3540
      ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 466 AspIleGluGluGlyGlyIleSerAlaPheSerArgSerGlu 479
      ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3541 GACGTAGATTAATGCTGTGACACACAAGCAAAATAGCAACATTTGAA 3582
      ||| ||| ||| ||| ||| ||| ||| ||| |||

```

Search completed: January 31, 2003, 07:22:41
 Job time : 132.079 secs

